

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:31.35 ; Search time 61.65 Seconds  
(without alignments)  
518,460 Million cell updates/sec

Title: US-09-147-955-6  
2407  
Perfect score: 1 MSRAHVLAATFPAGHINPA.....EEDGSSVNNLKVFIDEVYGI 461  
Sequence:

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues  
Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_12:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.potent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2407	100.0	461	10	Q9ZR25
2	1646.5	68.4	460	10	Q9ZR27
3	1583.5	65.8	443	10	Q9ZR26
4	1054.5	43.8	456	10	Q9ZR27
5	1052	43.7	467	10	P93709
6	1022.5	42.5	474	10	Q04930
7	1003.5	41.7	519	10	Q9ZVY2
8	988	41.0	455	10	Q9ZVY5
9	935.5	38.9	458	10	Q9ZVY5
10	779	32.4	449	10	Q9ZVY5
11	740	30.7	460	10	Q9ZVY5
12	730	30.3	449	10	Q9ZVY5
13	711.5	29.6	479	10	Q9ZVY5
14	697	29.0	456	10	Q9ZVY5
15	682	28.3	438	10	Q9ZVY5
16	682	28.3	438	10	Q9ZVY5
17	656	27.3	453	10	Q9ZVY5
18	631	26.2	475	10	Q9ZVY5
19	554	23.0	481	10	Q9ZVY5

Result No.	Score	Query Match	Length	DB ID	Description
20	535	22.2	482	10	Q9ZVY0
21	504.5	21.0	455	10	Q9ZVY0
22	485.5	20.2	495	10	Q9ZVY0
23	483	20.1	440	10	Q9ZVY0
24	479	19.9	462	10	Q9ZVY0
25	479	19.9	495	10	Q9ZVY0
26	478	19.9	491	10	Q9ZVY0
27	475.5	19.8	496	10	Q9ZVY0
28	471.5	19.6	496	10	Q9ZVY0
29	466.5	19.4	484	10	Q9ZVY0
30	461.5	19.2	496	10	Q9ZVY0
31	457	19.0	476	10	Q9ZVY0
32	451.5	18.8	454	10	Q9ZVY0
33	447	18.6	477	10	Q9ZVY0
34	445	18.5	476	10	Q9ZVY0
35	431	17.9	347	10	Q9ZVY0
36	429	17.8	466	10	Q9ZVY0
37	428	17.8	478	10	Q9ZVY0
38	425	17.7	481	10	Q9ZVY0
39	424.5	17.6	474	10	Q9ZVY0
40	418	17.4	455	10	Q9ZVY0
41	410	17.0	460	10	Q9ZVY0
42	405.5	16.8	488	10	Q9ZVY0
43	403.5	16.8	447	10	Q9ZVY0
44	397	16.5	456	10	Q9ZVY0
45	394	16.4	452	10	Q9ZVY0

## ALIGNMENTS

RESULT 1  
ID Q9ZR25 PRELIMINARY; PRT; 461 AA.  
AC Q9ZR25;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE.  
GN HGT8.  
OS Verbena x hybrida.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Asteridae; euasterids I; Lamiales; Verbenaceae;  
OC Verbenaceae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PETAL;  
RX MEDLINE: 99167509.  
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,  
RA KOSUMI T., SAITO K.;  
RT "Molecular cloning and biochemical characterization of a novel  
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for  
RT plant forms regarding anthocyanin".  
RL J. Biol. Chem. 274:7405-7411(1999).  
DR EMBL: AB013598; BAA36423.1; -;  
DR PROSITE: PS00375; UDPGT; 1.  
KW TRANSFERASE.  
SQ SEQUENCE 461 AA; 51347 MW; 6DCBC3B6 CRC32;

Query Match 100.0%; Score 2407; DB 10; Length 461;  
Best Local Similarity 100.0%; Pred. No. 3e-178;  
Matches 461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	DB	Sequence
QY	1	MSRAHVLAATFPAGHINPA.....EEDGSSVNNLKVFIDEVYGI 461
DB	1	MSRAHVLAATFPAGHINPA.....EEDGSSVNNLKVFIDEVYGI 461
QY	61	SFSDGYDDGLPDDGDKNYSKRSKIKALSTLIANNVDOSSKITYVYSHLFMAA 120
DB	61	SFSDGYDDGLPDDGDKNYSKRSKIKALSTLIANNVDOSSKITYVYSHLFMAA 120

QY 238 AFLGKDPDSGSDGDLFEKSGNDDCJLEWJSTNPRSSVYVFCFSFVNTKRSQMEETAR 297  
 Db 234 AFLDGGDPSSEISYGGDLFEK--SEENKCYEMLDTPRKSSVYVVSQSVLRPFKAQMEETGK 292  
 QY 298 GLDLCGRPFLWVVR-----NEGEVLISCMEEELKRVGKIYVSCSQLEVLTHPSIGCFVT 352  
 Db 293 GLAACGRPFLWIRQKNDGDEEBEELSCIGELKMKGIYVSCQLEVLALHPALGCFVT 352  
 QY 353 HCGNASTLESISFGVPMVAFQWPEQGCINAKLMDEVMFTGVRVVRANEGSVYDGDDEIRRC 412  
 Db 353 HCGNMSAVESLSCGPVVAVPWPEQDTTNAKLIDEMWGTGVRVVRANEGG--VDGSEIBRC 411  
 QY 413 IEEVMDGGEKRLRESAGKKNDLARKAMEEDSSVNNKLFYDEL 458  
 Db 412 VEMVMDGGEKSLVRENMIKKTLAREMGDGSGLKLNLFHY 457  
 RESULT 3  
 Q9ZR26 09ZR26 PRELIMINARY; PRT; 443 AA.  
 AC 09ZR26  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTTRANSFERASE HOMOLOGUE.  
 GN PF3R6.  
 OS Perilla frutescens.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Asteridae; eusterids I; Lamiales; Lamiales; Perilla.  
 RN RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, SHIKON; TISSUE-LEAF;  
 RX MEDLINE; 99167509.  
 RA YAMAZAKI M., GONG Z., FOKUCHI-MITUTANI M., FUKUI Y., TANAKA Y.,  
 RA KUSUMI T., SAITO K.;  
 RT "Molecular cloning and biochemical characterization of a novel  
 RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for  
 RT plant forms regarding anthocyanin."  
 RL J. Biol. Chem. 274.7405-7411(1999).  
 DR EMBL; AB013597; BAA36422.1; -.  
 KW PROSITE; PS00375; UDPGT; 1.  
 KW transferase.  
 SQ SEQUENCE 443 AA; 49110 MM; D87B5486 CRC32;  
 Query Match 65.8%; Score 1583.5; DB 10; Length 443;  
 Best Local Similarity 68.7%; Pred. NO. 1,4e-114;  
 Matches 310; Conservative 57; Mismatches 65; Indels 19; Gaps 7;  
 QY 1 MSRAHVLLATPPAOGHINPALOFARLNNADIQVFTSVYARRMRST---AAGSNGLI 57  
 Db 1 MVRRAVLLATPPAOGHINPALDFARLKLACGIDVFTFSVYARRMANTASAAAGNPGGL 60  
 QY 58 NFVFSFDYDGLDGLDGGDKNYMSEMKSRGKALSDTLAANNVDKSSKIFVYVSHLFA 117  
 Db 61 DFVAASDGYDGLDKRGKGGKRYMSEMKARGSALNN-LLLNND-----VFVYVSHLFA 114  
 QY 118 WAKYARFELRLSALIMLEPATVLDIFFYFNGYSDEIDAGSDAHLHGGGLPYLAQRDLR 177  
 Db 115 WAAEVARLSHPVATLAWPEPAIVLCIYHFYNGYADEIDAGSNEIOLP-RLPSLRQSLRP 173  
 QY 178 SFLPSLTERHFRPLSKKLETLTEGEERKRVVNSDADLEPDLKALIDKYMIAIGPLIPS 237  
 Db 174 TFLPLATERRRLMKKEKLETLIDGEEKKAVLVNITDLEPDLTLIDREKLIGIGPLIPS 233  
 QY 238 AFLDCKDPSDRSFGDLEFEKSGNDDCJLEWJSTNPRSSVYVVSFGSFVNTKRSQMEETAR 297  
 Db 234 AFLDGEDPSSEISYGGDLFEK--SEENKCYEMLNKPKSSVYVVSFGSVLRFPAQMEETGK 292  
 QY 298 GLDLCGRPFLWVVR-----VNEGEVLISCMEEELKRVGKIYVSCSQLEVLTHPSIGCF 350  
 Db 293 GLAACGRPFLWIRQKNDGDEEBEELSCIGELKMKGIYVSCQLEVLALHPALGCFVT 352



Db 293 GILACRPFLMIRKNDGEEEEEELSCIGELKKMKIVSMCSOLEVLAHPALGCF 352  
Qy 351 VTHCGMNSTLESISFGVPVAPQPFQDGTNAKLMEDVNTGVRVRANEGSVYDDEIR 410  
Db 353 VTHCGMNSTLESISFGVPVAPQPFQDGTNAKLMEDVNTGVRVRANEGSVYDDEIR 411  
Qy 411 RCIEEYMDGEGSRKLRBSAGKWKDLARKAM 441  
Db 412 RCIEEYMDGEGSRKLRBSAGKWKDLARKAM 442  
RESULT 4  
023270 PRELIMINARY: PRT: 456 AA.  
AC 023270:  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE GLUCOSYLTRANSFERASE.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE; 98121113.  
RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,  
RA BERGMAN R., DIRSE W., VAN STAVEREN M., STEKEMA W., DROST L.,  
RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIERFANELLI P., WEDDER H.,  
RA WEDDER H., MAMBERT R., WEITZINGER T., POHL T.M., TERRY N.,  
RA GIENEN J., VILLARUEL R., DE CLERCK R., VAN MONTAGI M., LECHARNY A.,  
RA AUDRONS S., CY I., KREIS M., LHO N., KAVANAGH T., HEMPEL S., KOTTER P.,  
RA ENTIAN K.D., RIBER M., SCHAEFER W., FUNK B., MUELLER-AUER S.,  
RA SILVER M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,  
RA VOURELATO E., MILLIONI D., HATZIOPOULOS P., PIRAVANDI E., OBERMAIER B.,  
RA HILBERT H., DUESTERHOEF A., MOORES T., JONES J.D.G., ENEVA T.,  
RA PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,  
RA DEUSEN M., VOET M., VOLCKAERT G., MEMES H.W., KLOSTERMAN S.,  
RA SCHUELLER C., CHALWATZIS N.,  
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of  
RT Arabidopsis thaliana."  
RL Nature 391:485-488(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z97335; CAB10189.1;  
DR PROSITE; PS00375; UDPGT; 1.  
DR PFM; PF00201; UDPGT; 1.  
RW Transferase.  
SO SEQUENCE 456 AA; 50702 MW; EAE0D380 CRC32;

Qy 235 IPSAFIDGKDPBDRSGDLFEKGSNDGCLLEMLSTNPSSVYVYSGFV-NTTSQME 293  
Db 241 VSSS--EGRT-----DLFK--SSDEYTKMLDSEKSEYIYISLGHADLPKRNH 288  
Qy 294 EIRAGLLDGRPLVAVVRNNEGEEVLISCMEEELK--VGKIVSMCSOLEVLAHPALGCF 350  
Db 289 ALTHGVLANNRPFLLVIREKNPEKKRNFLERISDRGLVYVGSQYAVLAHCAVGC 348  
Qy 351 VTHCGMNSTLESISFGVPVAPQPFQDGTNAKLMEDVNTGVRVRANEGSVYDDEIR 410  
Db 349 VTHCGMNSTLESISFGVPVAPQPFQDGTNAKLMEDVNTGVRVRANEGSVYDDEIR 407  
Qy 411 RCIEEYMDGEGSRKLRBSAGKWKDLARKAMEDGSSVNNKAFVDE 457  
Db 408 RCIEEYMDGEGSRKLRBSAGKWKDLARKAMEDGSSVNNKAFVDE 454  
RESULT 5  
ID P93709 PRELIMINARY: PRT: 467 AA.  
AC P93709:  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE GLUCOSYLTRANSFERASE.  
GN UGT.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;  
OC Nicotiana.  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-BY-2; TISSUE-SUSPENSION CULTURE;  
RC KOJIMA H., HASHIZUME K., IMANISHI S., NAKAMURA K.;  
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB006623; BA19155.1;  
DR MENDEL; 9421; Nicta; 1525; 9421.  
DR PROSITE; PS00375; UDPGT; 1.  
DR PFM; PF00201; UDPGT; 1.  
RW Transferase.  
SO SEQUENCE 467 AA; 52989 MW; 3726B31 CRC32;

Query Match 43.7%; Score 1052; DB 10; Length 467;  
Best Local Similarity 45.1%; Pred. No. 1.9e-73;  
Matches 211; Conservative 93; Mismatches 140; Indels 24; Gaps 11;

Qy 2 SRAHVLATFPAGHINPALQFAKRLANDIOVTFSTVYAMRMSRTAAGSNGLINFVS 61  
Db 6 NECHVIALFPGGHINPILQFSGKRLINLGVKTLSSLSAFNRI-KNLPKIEGL-TFAP 63  
Qy 62 FSDGYDDGLOPG--DQGNMSEKSKGICALSSTLANVNDOKSR---TFVYSHLTA 117  
Db 64 FSDGYDGNKSGFDDYHLNLSATKSGSEFT-----ANLKSRAKNGYPTFRIVYITLMD 118  
Qy 118 MAKVAREFHLSALIMIPATVLDIFYFYNGYSD---EIDGSDAHLPGCLPLAOR 174  
Db 119 WAGSVAKKHLIPSTLFWIQPATVLDIFYFYNGYSD---EIDGSDAHLPGCLPLAOR 177  
Qy 175 DLPSFLPSTHERFSL--MKELTELEGEERKPVLVNSFDALPDLAKAIDKYMIAIG 232  
Db 178 DEPSFVFDVKSNDMAVESIKRQIEILNDEENRILVNTFDELALNLRKLVNTVAVGIC 237  
Qy 233 PLIPSAFLDGKDPBDRSGDLFEKGSNDGCLLEMLSTNPSSVYVYSGFVNTTSQME 292  
Db 238 PLIPSAFLDGKDPBDRSGDLFEKGSNDGCLLEMLSTNPSSVYVYSGFVNTTSQME 294  
Qy 293 EIRAGLLDGRPLVAVVRNNEGEEVLISCMEEELKRGKIVSMCSOLEVLAHPALGCF 349  
Db 295 EIRAGLLDGRPLVAVVRNNEGEEVLISCMEEELKRGKIVSMCSOLEVLAHPALGCF 354

```

OY 350 FVTHCGNNTLESISFCVPMVAFPOWFDGNTAKLMEDVMTGVRANEGSVVDEI 409
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 355 FLTHCGNNTLESISFCVPMVAFPOWFDGNTAKLMEDVMTGVRANEGSVVDEI 413
OY 410 RCTEEVMDGKSRKLRBSAGKKDLARKAMEGSSVNNKLVFDE 457
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 414 OKCEIYMGDAEEGEELRNKAKMDLAKESTRENNSSNNLAKAYNE 461

RESULT 6
ID 004930 PRELIMINARY; PRT; 474 AA.
AC 004930;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE UDP-GLUCOSE:INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE
DE (EC 2.4.1.121) (INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE).
GN IAGLU.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
OC core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae:
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA GILAD A., SCOLNIK P.A., BAR-ZVI D.;
RL Plant Physiol. 113:1004-1004(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA THORNBURG R.W., GRAHAM R.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1 CATALYTIC ACTIVITY: UDP-GLUCOSE + INDOLE-3-ACETATE -> UDP +
INDOLE-3-ACETYL-BETA-1-D-GLUCOSE.
EMBL: U81293; AAB58497.1; -
DR MENDEL; 16583; Arabid; 2543; 16583.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 474 AA; 53740 MW; 35521D70 CRC32;

Query Match 42.5%; Score 1022.5; DB 10; Length 474;
Best Local Similarity 45.8%; Pred. No. 3.6e-71;
Matches 221; Conservative 81; Mismatches 128; Indels 53; Gaps 16;

OY 5 HVLLATPPAGCHINPALQPAKRLN--ADIQVTFSTVYAM--RMSSTAAGSNLIYVS 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 13 HFLVTFPPAGCHINPDLAKRLAGTISGARVTPAASISAVNRRMFSTENVPETLI-FAT 71
OY 62 FSDGYDGLQPG-----DDGKNYSEMKSRGKALSDTLAANNVDKSKITFVYVS 113
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 72 YSDHDGDFKSSAYSKSDQDATGNFSEMRKRKKEITLIEIDNR--KQNRPTCYVYT 129
OY 114 HLFMAAKVAREFHLSALLMIEPATVLDIFFYFYNGYSDEI---DAGSDAHLPGGLP 169
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 ILLTWVAELAEFHLPSALLMVGQVTFVSIHYFYNGYEDAISEMANTPSSIKLP-SLP 188
OY 170 VLAORLPFLPSTHERF--RSLMKKELETLEGEKRVLYNSFDALPALAKI-DKYE 227
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 189 LLTVRDIPSPFVSSNVYAFLLPARQIDSLKEINPKILNTPQLEPEAMSSVPNFK 248
OY 228 MIATGPIIPSAFLDGKDPDRSFGDLPFEKGSNDQCLEMLSTNPRSSVYVSGSFVNT 287
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 IVPVGPL-----TLRTDPSRK---EYELMDTADSSVLVSFTLAVL 291
OY 288 TKSMEEIARGLDCGRPLVVR-----VNEGEVLLISCMEE-LKRVGKIVSMCS 337
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 292 SKQLVLELCAKLIQSRPFLMVTDKSVRNKDEQKEEDCISSEKSPDEIGVVMSCD 351
OY 338 QLEVLHPISLGEVTHCGNNTLESISFCVPMVAFPOWFDGNTAKLMEDVMTGVRV-- 395
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 352 QFRVLNHRISICFVTHCGNNTLESISFCVPMVAFPOWFDGNTAKLMEDVMTGVRVME 411

```

```

OY 396 RANEG-SVVDGEIRRCIEEVMGKSRKLRBSAGKKDLARKAMEGSSVNNKLVF 454
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 412 KKEEGVYVVDSEIRRCIEEYME--DKAEFFRNATRMKDLAEAVREGSSFNHLKAF 469
OY 455 LDE 457
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 470 VDE 472

RESULT 7
ID 092VY2 PRELIMINARY; PRT; 519 AA.
AC 092VY2;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE T25N20.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
OC core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae:
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA SHINN P., BUEHLER E., DEMAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONNAY A., CONNAY A., KURTZ D., OUI O., SHEN Y.K., TORIUMI M.,
RA VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T25N20."
RL EMBL: AC005106; AAC80599.1; -
DR PROSITE; PS00375; UDPGT; 1.
DR TRANSFERRASE; 519 AA; 58584 MW; C982007F CRC32;

Query Match 41.7%; Score 1003.5; DB 10; Length 519;
Best Local Similarity 46.2%; Pred. No. 1.2e-69;
Matches 222; Conservative 78; Mismatches 132; Indels 49; Gaps 15;

OY 1 MSRAVLLATPPAGCHINPALQPAKRL-ANADIQVTFSTVYAMRRSRRAAGSGL--I 57
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 51 MAPHFLVTFPPAGCHINPDLAKRLAGTISGARVTPAASISAVNRRMFSTENVPETLI-FAT 71
OY 58 NFVFSDDGYD--GLQPDGDKNYSEMKSRGKALSDTLAANNVDKSKITFVYVS 115
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 109 SFLFSDGFDGGSITVEDKRSVNLKYNOKDALSPFIETKRGD---SPVCLITITL 165
OY 116 FAMAQVAREFHLSALLMIEPATVLDIFFYFYNGYSDEIDAGSDAHLPGGLPVLAAQD 175
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 166 LMAPKVAAREFHLSALLMVGQVTFVSIHYFYNGYEDAISEMANTPSSIKLP-SLP 188
OY 176 LPSFLLS--THERRSIMKELETLEGEKRVLYNSFDALPALAKI-DKYEIAIGPL 234
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 218 LPSFLTPSNTKGAYDAFOEMMEFLIKETPKILINFDLSLEPALFAFPIDVAAVAGPL 277
OY 235 IPSAFLDG---KDPDSFGDLPFEKGSNDQCLEMLSTNPRSSVYVSGSFVNTKSO 291
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 278 LPTIFGSGTNSKXKDDSSITL-----WLDKSTESSVITYVSGTINELSKKO 325
OY 292 MEELIARGLDCGRPLVVR-----RVNBE-----EVLISCMEELEKRVKIVSMCSQ 339
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 326 IEELIARGLDCGRPLVVR-----RVNBE-----EVLISCMEELEKRVKIVSMCSQ 339
OY 340 EVLTHPSLGEVTHCGNNTLESISFCVPMVAFPOWFDGNTAKLMEDVMTGVRVANE 399
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 386 EVLSHRAVGEVTHCGNNTLESISFCVPMVAFPOWFDGNTAKLMEDVMTGVRVANE 445
OY 400 EGSVYVDDDEIRRCIEEVMGKSRKLRBSAGKKDLARKAMEGSSVNNKLVFLEDEV 459
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 446 DG-LVERGEIRRCLEAVE--EKSVELRENAKKKRLAMEAGREGSSDKMEAFVIEDIC 502
OY 460 G 460

```

Db 503 G 503

RESULT 8  
092ZY5 PRELIMINARY; PRT; 455 AA.

AC 092ZY5  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE T25N20.17.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;  
OC eukaryophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SHINN P., BUEHLER E., DEMAR K., PENG J., KIM C., LI Y., SUN H.,  
RA CONWAY A., CONWAY A., KURTZ D., OUI O., SHEN Y.K., TORIUMI M.,  
RA VYOTSAKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,  
RA ECKER J.R.;  
RT \*Genomic sequence for Arabidopsis thaliana BAC T25N20.\*;  
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AC005106; AAC80596.1;  
DR PROSITE: PS00375; UDEGT; 1.  
SQ SEQUENCE 455 AA; 51190 MW; F9BAC33E CRC32;

Query Match 41.0%; Score 988; DB 10; Length 455;  
Best Local Similarity 45.3%; Pred. No. 1.6e-68;  
Matches 213; Conservative 84; Mismatches 139; Indels 34; Gaps 12;

QY 1 MSRAHVLATPFAOGHINPALQFAKRL-ANADIQVTFSTSVYARMSRTAAGSGLINF 59  
DB 1 MAQPHFLVTFPAOGHIVNSLSPARLRITKGARVTFACLSVYHRSMPINHNVENLSP 60  
QY 60 VFSFDGIDYG-LQPGDDGNMYSEKSKGKIKALSDTLAANNVDKSSKITFYVYSHLPM 118  
DB 61 LFFSDGFDGIVISNTDDVONRLVHFERNKDALSDFIETAN--ONGDSPVSCLYITLIPMW 118  
QY 119 AAKVAREFLALMTEPATVLDIFYFYFNGYSDSIDAGSDAHLPGSLPYAOGDLP 178  
DB 119 VKVARREFLPBVHMLQPAFADITYN-----STGNNSVFEPPLPSLEIDLP 170  
QY 179 FLPLP-THERFSLMEKEKLETEGEEKPKVYNSFDALPDALKAIDKYMIAIGPLIPS 237  
DB 171 FLSPSTNNAAQAVYOELMDFLKEBSPKILVNTFDSLEPEFLTALPNIEMAVGFLTA 230  
QY 238 AFLDGKPSDRSFGDLEFKSGSNDDCLEMLSTNPRSSVYVYVSGSFVNTTNSOMEIAR 297  
DB 231 EIFTSES-----GKDLSDRHQSSSYTL-WLDSKTESSVIYVSGFMVELSKQIIEELAR 284  
QY 298 GLLDGRPLMAYV--RYN-----EGE-----EVLISCMELKRVKIVGMSQLELTPR 345  
DB 285 ALIEGGRPLMAYITDLINEAKIEGEETEIEKIAIGFRHELEVGATVWCQIDVLRHR 344  
QY 346 SLGCEVTHCGMNSTLESIFGVPYAFPMQPDGTNAKLMEVDWPRGVRARANEESVVD 405  
DB 345 AIGCGLTHCGMSSLESILVGPVYAFPMQSDQPAKAKLEELTKRGVAVRENSG-LVE 403  
QY 406 GDEIRRCIEVDGGEKSKRLRESAGKDLARKAMEEDGSSVNNLKAVL 455  
DB 404 RGEIRKCLEAVMEA--KSVLEIRENAKMKRLATAGREGGSSDKNEAVY 451

RESULT 9  
023406 PRELIMINARY; PRT; 458 AA.

AC 023406  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE SIMILAR TO INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;  
OC eukaryophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BEVAN M., STICKENA W., MURPHY G., WAMBUTT R., POHL T., TERRYN N.,  
RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,  
RA PUIGDOMENECH P., HATZOPOULOS P., OBERHAER B., DUESTERHOFT A.,  
RA JONES J., PALME K., ANSORGE W., DEISENY M., BANCROFT I., MEMES H.W.,  
RA SCHUELLER C., CHALMARTZIS N.;  
RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;  
RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.  
DR EMBL: Z97339; CAB10333.1;  
DR MENDEL: 26727; Arabid.2543;26727.  
DR PROSITE: PS00375; UDEGT; 1.  
DR PFAM: PF00201; UDEGT; 1.  
DR Transferrase.  
KW Submitted.  
SQ SEQUENCE 458 AA; 52005 MW; 67943A2F CRC32;

Query Match 38.9%; Score 935.5; DB 10; Length 458;  
Best Local Similarity 42.7%; Pred. No. 1.8e-64;  
Matches 206; Conservative 82; Mismatches 126; Indels 69; Gaps 16;

QY 5 HYLAIFPPAOGHINPALQFAKRLAN--ADIQVTFSTSVYARMSRTAAGSGLINF 59  
DB 13 HFLVTFPPAOGHINPLSLAKRLAGTISGARVTFASISAYNRMRSTENVPTLI-FAT 71  
QY 62 FSDGYDDGLOPG-----DDGNMYSEKSKGKIKALSDTLAANNVDKSSKITFYVYSHLPM 113  
DB 72 YSDGHDDGFRSSAYSDKSRDADATGNFMSKRRKKEKLETLIEDNR--KONRFTCYVIT 129  
QY 114 HLFMAKAVAREFLHRSALLMTEPATVLDIFYFYFNGYSDSI---DAGSDAHLPGSLP 169  
DB 130 ILTVAELA-----LFSIFYFYFNGYSDSISEMANTPSSIKLP-SLP 172  
QY 170 VLAQRLPFLPSTHERF-RSLMEKEKLETEGEEKPKVYNSFDALPDALKAIDKYMIAIGPLIPS 237  
DB 173 LTVRPDIPSEIVSSNYAFALPFAFREQIDSLKEEINPKILINFTFOLEPEAMSSVVDNFR 232  
QY 228 MIAIGPLISAFIDGKPSDRSFGDLEFKSGSNDDCLEMLSTNPRSSVYVYVSGSFVNTTNSOMEIAR 297  
DB 233 IYVGGPL-----TLRTPSSRG---ETIEMLDTRKADSSVLYVSGFTLAVL 275  
QY 288 TKSOMEIARGLLDCGRPLMAYV-----RVNDEGEVLISCMELKRVKIVGMSQLELTPR 345  
DB 276 SKQLVELICKALQSRPLMAYITDLSYRNKKEDEKEDCISFRELDELIGMNVSWCD 335  
QY 338 QLEVLTHPSLGCYTHCGMNSTLESIFGVPYAFPMQPDGTNAKLMEVDWPRGVRARANEESVVD 405  
DB 336 QFRVNLHRSIGCFVTHCGMNSTLESIVSGVYAFPMQDQMANAKLEELTKRGVAVRENSG-LVE 403  
QY 396 RANEEB-SVVDGDEIRRCIEVDGGEKSKRLRESAGKDLARKAMEEDGSSVNNLKAVL 454  
DB 396 KKEEBGVVYVDESEIRRCIEVME--DKAEFGNATRWKDLAAEAVREGGSSFNHLKAF 453  
QY 455 LDE 457  
DB 454 VDE 456

RESULT 10  
022822 PRELIMINARY; PRT; 449 AA.

AC 022822  
DT 01-JAN-1998 (TREMBlrel. 05, Created)

[illegible]

Query Match	29.0%;	Score 697;	DB 10;	Length 456;
-------------	--------	------------	--------	-------------



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:50:21 ; Search time 47.63 Seconds

(without alignments)  
229.252 Million cell updates/sec

Title: US-09-147-955-6

Perfect score: 2407

Sequence: 1 MSRAHVLAFFPAGCHINPA.....EDGSSVNLKVFPEVVG1 461

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2407	100.0	461	1	W092949
2	1646.5	68.4	460	1	W092947
3	1583.5	65.8	443	1	W092948
4	1473.5	61.2	478	1	W092950
5	1320	54.8	468	1	W092952
6	773.5	32.1	466	1	W09825
7	751.5	31.2	471	1	W03756
8	431	17.9	470	1	W47172
9	429.5	17.8	433	1	W13406
10	425	17.7	471	1	W98009
11	405.5	16.8	488	1	W64569
12	377	15.7	473	1	W56451
13	340.5	14.1	471	1	R49245
14	147	6.1	515	1	W56750
15	140.5	5.8	506	1	R10429
16	136.5	5.7	414	1	R93982
17	135	5.6	530	1	W47126
18	125	5.2	533	1	R26153
19	125	5.2	534	1	R26154
20	103	4.3	74	1	R30165
21	102	4.2	473	1	W64392
22	102	4.2	1198	1	W64384
23	94.5	3.9	209	1	R85453
24	94	3.9	431	1	R07464
25	94	3.9	431	1	W87890
26	92	3.8	1974	1	W98391
27	91.5	3.8	491	1	R88093
28	90.5	3.8	434	1	R13493
29	90.5	3.8	479	1	R41942
30	88.5	3.7	365	1	W93141
31	88	3.6	460	1	R28827
32	86.5	3.6	460	1	W36114
33	86.5	3.6	800	1	W34479

34	85.5	3.6	1093	1	R42818	TMF. New protein C
35	84	3.5	816	1	R85870	WD-40 domain-contg
36	83	3.4	492	1	W34594	Archaeobacterium Ae
37	83	3.4	804	1	W22709	Leucyl-tRNA synthet
38	83	3.4	957	1	W20841	H. pylori transmem
39	83	3.4	1786	1	W24790	P. falciparum live
40	82.5	3.4	409	1	W71464	Glycosyltransferas
41	82.5	3.4	578	1	W10422	Pentillium chrys
42	82	3.4	417	1	W54358	Phosphoglycerate k
43	81.5	3.4	563	1	W56476	Amino acid sequenc
44	81.5	3.4	855	1	W01601	Aquifex pyrophilus
45	81	3.4	444	1	W90340	G. max truncated S

## ALIGNMENTS

RESULT 1  
ID W092949 standard: Protein: 461 AA.  
AC W092949;  
DT 14-MAY-1999 (first entry)  
DE W0905287 Seq ID 9.  
KM Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.  
OS Verbena hybrida.  
PN W0905287-A1.  
PD 04-FEB-1999.  
PE 16-JUL-1998; J03199.  
PR 25-JUL-1997; JP-200571.  
PA (SUNR) SUNTORY LTD.  
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
PI Yamazaki M;  
DR WPI; 99-142940/12.  
DR N-PSDB; X02828.  
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
used to transform plants for improvement of plant coloration  
PS Claim 2; Page 76-78; 89pp; Japanese.  
CC This invention describes the isolation of plant proteins which have  
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used  
CC in the production of plant varieties with improved colour, for example in  
CC the production of flower varieties with desired coloration and cut  
CC flowers harvested from them.  
SQ Sequence 461 AA;

Query Match 100.0%; Score 2407; DB 1; Length 461;  
Best Local Similarity 100.0%; Pred. No. 9.6e-250;  
Matches 461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSRAHVLAFFPAGCHINPALQFAKRLANDIOVTFSTYAMRRMRTAAGSGLINPV	60
DB	1	MSRAHVLAFFPAGCHINPALQFAKRLANDIOVTFSTYAMRRMRTAAGSGLINPV	60
QY	61	SFSDGYDGLQPPGDDGNNYSEMKSRGIRKALSDTLAANNVDKSSKITFVYVSHLFMAA	120
DB	61	SFSDGYDGLQPPGDDGNNYSEMKSRGIRKALSDTLAANNVDKSSKITFVYVSHLFMAA	120
QY	121	KVAREFLRSALMIWEPATVLDIFYFYNGYSDEIDAGSDAHLPGILPYLAORDPSFL	180
DB	121	KVAREFLRSALMIWEPATVLDIFYFYNGYSDEIDAGSDAHLPGILPYLAORDPSFL	180
QY	181	LPSTHERFRSLMKRKLTLTGEEKPKVLVNSFDALPDALKAIDKYMAIGPLIPSAFL	240
DB	181	LPSTHERFRSLMKRKLTLTGEEKPKVLVNSFDALPDALKAIDKYMAIGPLIPSAFL	240
QY	241	DGKDPDRSFGDLEFGSNDQCLEWLTNPSSVYVYFSGSVVNTTKSOMERIAGTL	300
DB	241	DGKDPDRSFGDLEFGSNDQCLEWLTNPSSVYVYFSGSVVNTTKSOMERIAGTL	300
QY	301	DCGRPFLLVVRVNGEVEVLISCMEEELKRVGIYSWCQLFVLTHTPSLIGCVTHCGMNSTL	360
DB	301	DCGRPFLLVVRVNGEVEVLISCMEEELKRVGIYSWCQLFVLTHTPSLIGCVTHCGMNSTL	360

QY 361 ESISFGVPMVAFPMQFDGCTNAKLMEVDWRTGVVRANEGSSVVDGDEIRRCIEVDGG 420  
DB 361 ESISFGVPMVAFPMQFDGCTNAKLMEVDWRTGVVRANEGSSVVDGDEIRRCIEVDGG 420  
QY 421 EKSRLRESAGKMKDLARKAMEEDGSSVNNLKVFLEVDYGI 461  
DB 421 EKSRLRESAGKMKDLARKAMEEDGSSVNNLKVFLEVDYGI 461

## RESULT 2

W92947  
ID W92947 standard; Protein; 460 AA.  
AC W92947;  
DE 14-MAY-1999 (first entry)  
DE W0905287 Seq ID 7.  
KM Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.  
OS Perilla frutescens.  
PN W0905287-A1.  
PD 04-FEB-1999.  
PE 16-JUL-1998; J03199.  
PR 25-JUL-1997; JP-200571.  
PA (SUNR ) SUNTORY LTD.  
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
PI Yamazaki M;  
DR WPI; 99-142940/12.  
DR N-PSDB; X02826.  
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
used to transform plants for improvement of plant coloration  
PS Claim 2; Page 71-73; 89pp; Japanese.  
CC This invention describes the isolation of plant proteins which have  
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used  
CC in the production of plant varieties with improved colour, for example in  
CC the production of flower varieties with desired coloration and cut  
CC flowers harvested from them.  
SQ Sequence 460 AA;

Query Match 68.4%; Score 1646.5; DB 1; Length 460;  
Best Local Similarity 70.0%; Pred. No. 4.4e-168;  
Matches 326; Conservative 49; Mismatches 74; Indels 17; Gaps 7;

QY 1 MSRAHVLATTPAGCHINPALQFAKRLANDIOVTFFTSVYAMRMKST---ANGSGLI 57  
DB 1 WVRRRVLTATTPAGCHINPALQFAKRLACTDVTFFTSVYAMRMKNTASAAAGNPGCL 60  
QY 58 NFVFSFGYDGDGLPGDDGKNYSEMSRGIKALSDTLAANNVQKSKITFEVYSHLFA 117  
DB 61 DFVAFSDGIDGDLKPGDGRKYNSEMKARSGEALRN-LLNNHD----VTFVYSHLFA 114  
QY 118 WAAKVAEFLHLSALMLIEPATVLDIFYFYNGYSDETDAGSDAHLPGGLVLAQRDL 177  
DB 115 WAAVARSQVPSALLWEPATVLCIYFYNGYADEIDAGSDEIQLP-RLPLEQSLP 173  
QY 178 SFLLPSTHERSRSLMKEKLETLGEDEKPKVLVNSFDALPDAIKDYEMAIAGPLPS 237  
DB 174 TELLPETPERRLMKKEKLETLGEDEKAKVLYNFDALPDAIDRYELIGIGPLPS 233  
QY 238 AFLDGKPSDSFGDGLFEKGSNDCCLEMLSTNPRSSVYVSGSFVNTTKSQMEELAR 297  
DB 234 AFLDGSDPSSEISYGDLEK-SEENNCVEMLDTPKSSVYVSGSVLRFPAKMEELGK 292  
QY 298 GLLDGGRPLWVVR-----NEGEEVLISCMELKRVGKIVSMCSQLEVLTHPSLGCFT 352  
DB 293 GLLAGGRPLMIMIREQKNDGDEEBEELISCIIGELAKMKKIVSMCSQLEVLTHPSLGCFT 352  
QY 353 HCGMNSTLESISFGVPMVAFPMQFDGCTNAKLMEVDWRTGVVRANEGSSVVDGDEIRRC 412  
DB 353 HCGMNSTLESISFGVPMVAFPMQFDGCTNAKLMEVDWRTGVVRANEGSSVVDGDEIRRC 412  
QY 413 IEEVMDGSEKSRKLRRESAGKMKDLARKAMEEDGSSVNNLKVFLEVDYGI 458  
DB 412 VEVMDGSEKSRKLRRESAGKMKDLARKAMEEDGSSVNNLKVFLEVDYGI 457

## RESULT 3

W92948  
ID W92948 standard; Protein; 443 AA.  
AC W92948;  
DE 14-MAY-1999 (first entry)  
DE W0905287 Seq ID 8.  
KM Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.  
OS Perilla frutescens.  
PN W0905287-A1.  
PD 04-FEB-1999.  
PE 16-JUL-1998; J03199.  
PR 25-JUL-1997; JP-200571.  
PA (SUNR ) SUNTORY LTD.  
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
PI Yamazaki M;  
DR WPI; 99-142940/12.  
DR N-PSDB; X02827.  
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
used to transform plants for improvement of plant coloration  
PS Claim 2; Page 74-76; 89pp; Japanese.  
CC This invention describes the isolation of plant proteins which have  
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used  
CC in the production of plant varieties with improved colour, for example in  
CC the production of flower varieties with desired coloration and cut  
CC flowers harvested from them.  
SQ Sequence 443 AA;

Query Match 65.8%; Score 1583.5; DB 1; Length 443;  
Best Local Similarity 68.7%; Pred. No. 2.4e-161;  
Matches 310; Conservative 57; Mismatches 65; Indels 19; Gaps 7;

QY 1 MSRAHVLATTPAGCHINPALQFAKRLANDIOVTFFTSVYAMRMKST---ANGSGLI 57  
DB 1 WVRRRVLTATTPAGCHINPALQFAKRLACTDVTFFTSVYAMRMKNTASAAAGNPGCL 60  
QY 58 NFVFSFGYDGDGLPGDDGKNYSEMSRGIKALSDTLAANNVQKSKITFEVYSHLFA 117  
DB 61 DFVAFSDGIDGDLKPGDGRKYNSEMKARSGEALRN-LLNNHD----VTFVYSHLFA 114  
QY 118 WAAKVAEFLHLSALMLIEPATVLDIFYFYNGYSDETDAGSDAHLPGGLVLAQRDL 177  
DB 115 WAAVARSQVPSALLWEPATVLCIYFYNGYADEIDAGSDEIQLP-RLPLEQSLP 173  
QY 178 SFLLPSTHERSRSLMKEKLETLGEDEKPKVLVNSFDALPDAIKDYEMAIAGPLPS 237  
DB 174 TELLPETPERRLMKKEKLETLGEDEKAKVLYNFDALPDAIDRYELIGIGPLPS 233  
QY 238 AFLDGKPSDSFGDGLFEKGSNDCCLEMLSTNPRSSVYVSGSFVNTTKSQMEELAR 297  
DB 234 AFLDGSDPSSEISYGDLEK-SEENNCVEMLDTPKSSVYVSGSVLRFPAKMEELGK 292  
QY 298 GLLDGGRPLWVVR-----VNEGEEVLISCMELKRVGKIVSMCSQLEVLTHPSLGCFT 350  
DB 293 GLLAGGRPLMIMIREQKNDGDEEBEELISCIIGELAKMKKIVSMCSQLEVLTHPSLGCFT 352  
QY 351 VTHGCMNSTLESISFGVPMVAFPMQFDGCTNAKLMEVDWRTGVVRANEGSSVVDGDEIR 410  
DB 353 VTHGCMNSTLESISFGVPMVAFPMQFDGCTNAKLMEVDWRTGVVRANEGSSVVDGDEIR 410  
QY 411 RCIEVMDGSEKSRKLRRESAGKMKDLARKAMEEDGSSVNNLKVFLEVDYGI 441  
DB 412 RCIEVMDGSEKSRKLRRESAGKMKDLARKAMEEDGSSVNNLKVFLEVDYGI 442

## RESULT 4

W92950  
ID W92950 standard; Protein; 478 AA.  
AC W92950;  
DE 14-MAY-1999 (first entry)  
DE W0905287 Seq ID 10.  
KM Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.



OS Torenla hybrida.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 64 /note= "Cys or Phe"  
 FT Misc\_difference 65 /note= "Ser or Pro"  
 FT Misc\_difference 66 /note= "Ser or Pro"  
 PN W09905287-A1.  
 PD 04-FEB-1999.  
 PE 16-JUL-1997; JP-200571.  
 PR 25-JUL-1997; JP-200571.  
 PA (SUNR ) SUNTORY LTD.  
 PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
 PI Yamazaki M;  
 PI WPI: 99-142940/12.  
 DR N-PSDB: X02829.  
 DR Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
 PT used to transform plants for improvement of plant coloration  
 PS Claim 2: Page 78-81; 89pp; Japanese.  
 CC This invention describes the isolation of plant proteins which have  
 CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used  
 CC in the production of plant varieties with improved colour, for example in  
 CC the production of flower varieties with desired coloration and cut  
 CC flowers harvested from them.  
 SQ Sequence 478 AA;

Query Match 61.2%; Score 1473.5; DB 1; Length 478;  
 Best Local Similarity 61.5%; Pred. No. 1.8e-149;  
 Matches 295; Conservative 71; Mismatches 85; Indels 29; Gaps 11;

QY 1 MSRAVYLATPPAOGHINPALOFAPKRLANADI--OVTEFTSYAMRMRTAAGSLIN 58  
 DB 2 VNKRIILLATPPAOGHINPSEFAKRLNTGTYDOVFTSYALRRM-RFETDSSRID 60  
 QY 59 FVFSFGYDDGLQPGDDGKNYSEKSRGKALSDTLAANNVQKSS--KITFVYSH 114  
 DB 61 FVAXXSDYDDGLKGGDDGKNYSEKSRGKALSDTLAANNVQKSS--KITFVYSH 120  
 QY 115 LFAAAKAAAREPHLSALIMTEPATVLIFFYFNGYSDEIDAGSAIHLPGLPYLAOR 174  
 DB 121 LFSMAAEAREVDPASALIMTEPATVLIFFYFNGYSDEIDAGSAIHLPGLPYLAOR 179  
 QY 175 DLPSEFLPSTHERFRL-----MKEKLEETLEGEERPKVLVNSFDALPDALEKADIKYEM 234  
 DB 180 DLPSEFLPSTHERFRL-----MKEKLEETLEGEERPKVLVNSFDALPDALEKADIKYEM 239  
 QY 235 IPSA-FLDGKDP--SDRSFGDLEFGKSNDDCLMFLSTNPRSSVYVYFSGFVNTTQSO 291  
 DB 240 IPSA-FLDGKDP--SDRSFGDLEFGKSNDDCLMFLSTNPRSSVYVYFSGFVNTTQSO 297  
 QY 292 MEELARGLDLCGRPLLVYRVN-----EGEYVLSCEEB--LKRVGKIVSWCSOLE 340  
 DB 298 MEELARGLDLCGRPLLVYRVN-----EGEYVLSCEEB--LKRVGKIVSWCSOLE 357  
 QY 341 VLTHTSLGCFVTHGCMNSTLESISFGVPVAPFQWPDQNTAKMEDVYRSGVRANEE 400  
 DB 348 VLTHTSLGCFVTHGCMNSTLESISFGVPVAPFQWPDQNTAKMEDVYRSGVRANEE 417  
 QY 401 GSVVGDDEIRRCIEEYVMDGGEKSRKLRSAKKMDLARKAMEED--GSSVNNLKVFLEDEV 459  
 DB 418 GSVVGDDEIRRCIEEYVMDGGEKSRKLRSAKKMDLARKAMEED--GSSVNNLKVFLEDEV 473  
 RESULT 5  
 ID W92952 standard; Protein: 468 AA.  
 AC W92952;  
 DT 14-MAY-1999 (first entry)  
 DE W09905287 Seq ID 12.  
 KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.  
 OS Petunia hybrida.  
 PN W09905287-A1.  
 PD 04-FEB-1999.

PF 16-JUL-1998; J03199.  
 PR 25-JUL-1997; JP-200571.  
 PA (SUNR ) SUNTORY LTD.  
 PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
 PI Yamazaki M;  
 PI WPI: 99-142940/12.  
 DR N-PSDB: X02831.  
 DR Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
 PT used to transform plants for improvement of plant coloration  
 PS Claim 2: Page 83-85; 89pp; Japanese.  
 CC This invention describes the isolation of plant proteins which have  
 CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used  
 CC in the production of plant varieties with improved colour, for example in  
 CC the production of flower varieties with desired coloration and cut  
 CC flowers harvested from them.  
 SQ Sequence 468 AA;

Query Match 54.8%; Score 1320; DB 1; Length 468;  
 Best Local Similarity 55.7%; Pred. No. 5.2e-133;  
 Matches 264; Conservative 76; Mismatches 108; Indels 26; Gaps 10;

QY 1 MSRAVYLATPPAOGHINPALOFAPKRLANADIQVTEFTSYAMRMRTAA--GSLIN 58  
 DB 1 MGPPIVLTTPAOGHINPALOFAPKRLANADIQVTEFTSYAMRMRTAA--GSLIN 59  
 QY 59 FVFSFGYDDGLQPGDDGKNYSEKSRGKALSDTLAANNVQKSSKITFVYSHLPM 118  
 DB 60 FVFSFGYDDGLQPGDDGKNYSEKSRGKALSDTLAANNVQKSSKITFVYSHLPM 117  
 QY 119 AAKVAAREPHLSALIMTEPATVLIFFYFNGY----SDEIDAGSAIHLPGLPYLAOR 174  
 DB 118 AAKVAAREPHLSALIMTEPATVLIFFYFNGY----SDEIDAGSAIHLPGLPYLAOR 176  
 QY 175 DLPSEFLPSTHERFRL-----MKEKLEETLEGEERPKVLVNSFDALPDALEKADIKYEM 228  
 DB 177 DLPSEFLPSTHERFRL-----MKEKLEETLEGEERPKVLVNSFDALPDALEKADIKYEM 234  
 QY 229 IATGPIIPSAFLDGKDPDSRSGDLEFGKSNDDCLMFLSTNPRSSVYVYFSGFVNTT 288  
 DB 235 IATGPIIPSAFLDGKDPDSRSGDLEFGKSNDDCLMFLSTNPRSSVYVYFSGFVNTT 291  
 QY 289 KSMEEIARGLDLCGRPLLVYRVN-----EGEYVLSCEEB--LKRVGKIVSWCSOLE 344  
 DB 292 KSMEEIARGLDLCGRPLLVYRVN-----EGEYVLSCEEB--LKRVGKIVSWCSOLE 351  
 QY 345 PSIGCFVTHGCMNSTLESISFGVPVAPFQWPDQNTAKMEDVYRSGVRANEEGSSVY 404  
 DB 352 PSIGCFVTHGCMNSTLESISFGVPVAPFQWPDQNTAKMEDVYRSGVRANEEGSSVY 410  
 QY 405 DQDEIRRCIEEYVMDGGEKSRKLRSAKKMDLARKAMEEDGSSVNNLKVFLEDEV 458  
 DB 411 DQDEIRRCIEEYVMDGGEKSRKLRSAKKMDLARKAMEEDGSSVNNLKVFLEDEV 464  
 RESULT 6  
 ID W09825 standard; Protein: 466 AA.  
 AC W09825;  
 DT 15-JUL-1997 (first entry)  
 DE UDP-glucose:thiohydroximate S-glucosyltransferase.  
 KW Glucosinolate; UDP-glucose:thiohydroximate S-glucosyltransferase;  
 OS Brassica napus cv. Westar.  
 FH Key Location/Qualifiers  
 FT misc\_difference 2 /note= "residue 2 is Val in other S-GT isoforms"  
 FT misc\_difference 10 /note= "a Lys residue is inserted between amino  
 FT misc\_difference 12 acids 10 and 11 in some S-GT isoforms"  
 FT misc\_difference 43 /note= "residue 12 is Ser in some S-GT isoforms"

Query Match	32.18;	Score	773.5;	DB 1;	Length	466;
Best Local Similarity	38.98;	Pred. No.	2.5e-74;			
Matches	102;	Conservative	77;	Mismatches	180;	Indels
					29;	Gaps
						12.

QY	2	SRAYVLATPEAQGHINPDAQFAKRLANMADQYTFPFSSVAAKRMSTAG-SNGLINPV	60
Db	13	SKGHVYVLPYQGHINPMQFAKRLVSKGVKVIATTV-----TASSISTPSVSE	65
QY	61	SFSDQYDGL--QCGDDCKNYSSEKSGIYALSDTLAANNVDQSSKIFPVVYSHLFAM	118
Db	66	PISDGHDFILGVGVSTIDASESFKLNGSETL--TVISKFKSTPDISDLYYDSFLP	123
QY	119	AAKVAREFHLRSALLMIEPATVLDIFYFYENGYSDEIDAGSD--AIHLPGLLPVLAQRD	175
Db	124	GLEVARSNSISAAAFNTNLTVCVSILKRFVSG---EPPLPADPASAPYLVGLPALSYDE	180
QY	176	LPSEFL--LPSTH--ERFRSLMEKLETLDEGEKPKVIVNSPDALRDLKAIIDYEMIA--	230
Db	181	LPSPFGHRSSSHAEGRVLL--QFRNHDEADMLFVNSFEGLETQCEGESEEMAKATL	237
QY	231	IGPLPSAFVLGDKPDSFSGDLEFKGSNDDCLEMLSTNPRESSVYVYFSGSFVMTKS	290
Db	238	IGPMPLPSAYLDGRKIDDKGVSSILMKPLPS--EECEMMLDTRKLSVYFVFSFGSFLIFK	295
QY	291	QMEIARGILDCGRPFLVAVVYVNGEEVYLISCMBELRVGKIVSWCSQLEVLTHPSLGF	350
Db	296	QAAEVAKALQSENFNEFLVAKIEAHIAKLPEGFVATKDRALVLVSMQGLEVLAVHSIGCF	355
QY	331	YTHGCMNSTLESISGVYMAVFPWFQDGTAKLMEDYVWRTGVAVRANEGSVYDGEIR	410
Db	336	LTHCGMNSTLEGLSLGVYMAVFPWMSQMDAKVVEEWVRYGVAKKEAGGVYKSDDEV	415
QY	411	RCIEVDGEGEKSRKLRSSACKWMDLARKAMEEFGSSVNNLKVPLEIV	458
Db	416	RCLRGVME-GESSVEIRRESSKKMDLAVKANSSECGSSDRSINEFVSL	462
RESULT	7		
ID	W03756		
AC	W03756	standard; protein; 471 AA.	
DT	27-OCT-1996	(first entry)	
DE	Maize UDP-glucose indol-3-yl:acetyl:glucosyl transferase.		
KM	UDP-glucose-indol-3-yl:acetyl:glucosyltransferase; enzyme;		
KW	transgenic plant; EC:2.4.1.121; crop improvement; corn; cereal;		
KW	grass; InaU-transferase.		
OS	Zea mays.		
FH	Key		
FT	region	Location/Qualifiers	
FT	region	1..279	
FT	region	/note="major hydrophilic region"	
FT	region	293..299	
FT	region	/note="major hydrophilic region"	
FT	region	399..406	
FT	region	/note="major hydrophilic region"	
FT	region	435..444	
FT	misc_difference	/note="major hydrophilic region"	
FT	misc_difference	363	
FT	misc_difference	/note="potential glycosylation site"	
FT	misc_difference	37	
FT	misc_difference	/note="potential protein-kinase-C phosphorylation site"	
FT	misc_difference	453	
FT	misc_difference	/note="potential protein-kinase-C phosphorylation site"	
FT	misc_difference	469	
FT	binding_site	/note="potential protein-kinase-C phosphorylation site"	
FT	binding_site	268..393	
FT	binding_site	/note="UDP-glucose binding site"	
PN	W0600291-AL		
PF	04-JAN-1996		
PR	20-JUN-1995; U07820.		
PR	24-JUN-1994; US-265427.		
PA	(RESE.) RESEARCH CORP TECHNOLOGIES INC.		
PI	Bandurski RS, Szczygiowski K, Szerszen JB;		
DR	WPI: 96-068875/07.		
DR	N-PSDB: T10106.		

**us-09-147-955-6.rag**

useful to develop products that alter signalling pathways in plants by altering of salicylic acid, jasmonic acid or ethylene (Claim 2; Fig 3; 52pp; English). This is a glucosyl transferase (GTrase) protein encoded by a wound inducible gene (TWII) isolated from wounded tomatoes. The TWII gene encodes this GTrase from amino acid position 5. The TWII gene can be used to identify homologue GTrase encoding genes isolated from tobacco and rice. A microbial host can be transfected or transformed with a vector containing the GTrase encoding nucleic acids. The products can be used to interfere with GTrase and therefore alter signalling pathways in plants, specifically tobacco, rice or tomato plants by altering levels of salicylic acid, jasmonic acid or ethylene. This can induce the production of plant defence proteins such as pathogenesis-related (PR) and proteinase inhibitor (PIN) proteins which regulate plant development (plant growth, reproduction and senescence) and improve plant response to

Query Match	17.9%	Score 431	DB 1	length 470
-------------	-------	-----------	------	------------

Matches 147; Conservative 68; Mismatches 200; Indels 94; Gaps 19;

QY 1 MSRAHVLLATFPAQGHINPALQFAKRLANADIQVTFET-----SVYAWRRMSRTAAGSNG 55

Db 1 MGELHFFFPDDAQGHMIP TLDMANV VACRGV KATII TTP LNESVF S KAIERNKHLGIEI 60

```

56 LINFVSFSDGYDD-----GLQPGDDG-KNYMSEMKSRGIKALSDTLAANNVDQKSSK 106
QY

```

DB 61 DIRLKEPAKENDLPEDCERDVLPSDDKLPNFK-----AAAMMKDEFEEL 10/

10/ 1 - - - 1 F V V I S H L E A W A A N V A N E F H U N S A L T W I E F F A I V U D I F I F I E N G I S D E I D A G S D A 101

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

[illegible]

00 307 --WVNSDPAI EBDIAKATDVVEMT-----AICDPIBDAEINCKVDBSDBSECCDI EFK 357

Db 211 SYGVTFNSFEYI.ESD---YVEHYTKVGRKNWATGPI.----SI<sup>C</sup>NRDIEDKAEFG---RK 260

QY 258 GSND DD-CLEW LSTN PRSS VVYVSFGSFVN TKTSOME E IARGLLD CGRPF LWWVRVNEGE 316

Db 261 SSIDEHACLKWLDSKSSSIIVYVCFGSTADFTTAQMQLAMGLEASGQDFIWIIRTGNE 320

QY 317 EVLISCMEELKRVGKIV-SWCSQLEVLTHPSLGCFTVTHCGWNSTLESISFGVPMVAFPQW 375

Db 321 WLPEGFERTKEKGLIRGWAQSVILDHEAIGAFVTHCGWNSTLEGISAGVPMVTWPF 380

3/6 FDQTNAKLMEBWRIGVR-----RANEESVBDGEIRRCIEVMDGGEKSRLRES 429

DD 301 AEQFNENLVIEVMKSGAGVGSADQMAKIASSE--VAKREAIANAKAIKRVN-ASEEIEGFKSK 43/

[illegible]

RESULT	9
W13A05	

ID	W13406	standard; 433 AA.
AC	W13406.	

DE Solanum melongena flavanoid-3-glucosyl transferase,  
DT 10-JUN-1997 (first entry)

KW pigment; production; seed; plant.  
KW Egyptian, flavanoid- $\beta$ -glucosyl, transferase; glucose; flavonoid;  
KW pigment; production; seed; plant.

PN J09056385-A.

PF 25-AUG-1995; 238943.







Db 394 GDOAYNTNKEEL 406

## RESULT 15

R10429

ID R10429 standard; Protein: 506 AA.

AC R10429:

DT 10-APR-1991 (first entry)

DE Ecdysteroid UDP-glucosyl transferase.

KM Ecdysteroid UDP-glucosyl transferase; EGT; insect control;

OS Autographa californica nuclear polyhedrosis virus.

PN MO9100014-A.

PD 10-JAN-1991

PF 29-JUN-1989; U03758.

PR 29-JUN-1989; US-373952.

PA (UYGE-) UNIV GEORGIA RES FO.

PI Miller LK, Orellilly D;

DR WPI: 91-036527/05.

DR N-PSDB: Q10336.

PT Biological insect control agents - comprising nuclear

PT polyhedrosis virus with inactivated EGT gene encoding

PT ecdysteroid UDP-glucosyl transferase

PS Disclosure: Page 14-14/1; 64pp; English.

CC The protein affects the growth, development or behaviour of insects.

CC The gene encoding the protein is either inactivated to prevent

CC feeding behaviour, inhibit growth and result in the earlier death

CC of the insect host.

CC Pref. the egt gene is inactivated by replacing a portion of the gene

CC with a bacterial sequence encoding beta-galactosidase (VEGTZ); or

CC part of the gene is deleted without replacement (VEGTDEL).

CC A (recombinant) ecdysteroid UDP-glucosyl transferase having this

CC amino acid sequence, or a sequence about 70% homologous,

CC is claimed (52-53).

SQ Sequence 506 AA;

Query Match 5.8%; Score 140.5; DB 1; Length 506;

Best Local Similarity 22.7%; Pred. No. 2.6e-06;

Matches 98; Conservative 68; Mismatches 146; Indels 119; Gaps 24;

```
QY 29 NADIVTFEFTSYVAMRRMR-----TAAGSNGLINFEVSFGYDDGLQPGDDGK 77
DB 73 NADMSVEQTKLVANSAMFRKRVSDTPTVTAANTLGLEW--FKQQFDN----- 121
QY 78 NTMSEKRSKIGALSDTLAANNVDKSSKITFEVYSHLFAMAKVAREFHLSALIMIEP 137
DB 122 -----INVRNLANNOTFPLVVEAFADYA--LVFGHLY-----DP 155
QY 138 ATVLDFIIFYENGYSDEID-AGSDAHLPGGLPVLAQRDLPSFLPSTHERFRSLMKEL 196
DB 156 APTVQIAPGY--GLAENFTVGAVARH-----PV-----HHPNIMRNFDDTEANVTE-- 202
QY 197 ETLBGEERKXVLVNSFDAL-----EPDALKADKXEMIAIG--PL-----IPSAPL 240
DB 203 --WRLYKFKILLANNSNALLKQOPGPPTIEKLNKVLQLLNLHPIDNNRPV----- 256
QY 241 DKRDPDSRSGDGLFEKGSNDDCLEWLS-----TNPSSVYVYSGFSVNTTKSQME 293
DB 257 ----PSVOYLGGGHIHVKSAP---LTKLSPVINAQMNKSGTIVYFSSSID-TKSFAN 308
QY 294 ETARGILLDCGR-----PLWVVVYVNEGEEVLISCMEELKRVGKIV--SWCSOLEVLTHPS 346
DB 309 EFLVYLINFKTLDMVTILMKI-----DDEV-----YKNITLPANVITQNMFNQRAVLRHKK 360
QY 347 LGCFTVHCWNTLSISGCVPMVAFPPQMFDOGTNAKLMEDEWRTGVRVRANEGSVYDG 406
DB 361 MAFTTQGGSLGSSDEALEGLIPVCLPMGDOFYHAKHKLQQL--GV-ARALDT-VTVSS 415
QY 407 DEIRRCIEEVM 417
DB 416 DQLLVAINDV 426
```

Search completed: August 1, 2000, 19:50:22  
Job time: 21502 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 1, 2000, 19:51:31 ; Search time 59.57 Seconds

(without alignments)  
118.629 Million cell updates/sec

Title: US-09-147-955-6

Perfect score: 2407

Sequence: 1 MSRAHYLAIFPAQGHINPA.....EDGSSVNNLKVFIDEVVG1 461

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	751.5	31.2	471	2	US-08-466-583-2
2	751.5	31.2	471	4	PCT-US95-07820-2
3	430	17.9	471	3	US-09-106-464-2
4	405.5	16.8	488	2	US-08-797-226-2
5	390	16.2	131	2	US-08-466-583-5
6	390	16.2	131	4	PCT-US95-07820-5
7	267	11.1	131	2	US-08-466-583-6
8	267	11.1	131	4	PCT-US95-07820-6
9	227	9.4	63	2	US-08-466-583-8
10	227	9.4	63	4	PCT-US95-07820-8
11	154.5	6.4	529	4	PCT-US92-00282-7
12	146	6.1	531	4	PCT-US92-00282-6
13	140.5	5.8	506	5	5180581-2
14	136.5	5.7	414	4	US-08-750-524-1
15	134	5.6	531	4	PCT-US92-00282-5
16	129	5.4	533	4	PCT-US92-00282-3
17	129	5.4	534	4	PCT-US92-00282-4
18	125	5.2	56	2	US-08-466-583-4
19	125	5.2	56	4	PCT-US95-07820-4
20	110	4.6	58	2	US-08-466-583-9
21	110	4.6	58	4	PCT-US95-07820-9
22	105	4.4	52	2	US-08-466-583-7
23	105	4.4	52	4	PCT-US95-07820-7
24	103	4.3	74	4	PCT-US92-00282-24
25	99.5	4.1	1258	1	US-08-310-912A-107
26	94	3.9	431	1	US-07-783-705A-2
27	91.5	3.8	491	1	US-08-489-733-3
28	91.5	3.8	491	2	US-08-993-581B-3

29	90.5	3.8	419	3	US-08-115-753-2	Sequence 2, Appl1
30	90.5	3.8	419	3	US-08-115-753-33	Sequence 33, Appl
31	88.5	3.7	419	2	US-08-282-197C-49	Sequence 49, Appl
32	88	3.7	365	1	US-08-186-833-4	Sequence 4, Appl1
33	85.5	3.6	1093	4	PCT-US93-03077-1	Sequence 1, Appl1
34	84	3.5	816	1	US-08-190-802A-54	Sequence 54, Appl
35	83	3.4	804	1	US-08-785-428-2	Sequence 2, Appl1
36	83	3.4	804	2	US-08-996-797-2	Sequence 2, Appl1
37	82.5	3.4	409	2	US-08-924-254-2	Sequence 2, Appl1
38	82.5	3.4	409	3	US-09-120-249-2	Sequence 2, Appl1
39	82	3.4	330	3	US-08-115-753-1	Sequence 1, Appl1
40	81.5	3.4	855	2	US-08-468-558-2	Sequence 2, Appl1
41	81	3.4	821	1	US-07-928-464-2	Sequence 2, Appl1
42	81	3.4	821	1	US-08-003-311B-2	Sequence 2, Appl1
43	81	3.4	821	1	US-08-261-432-2	Sequence 2, Appl1
44	81	3.4	821	4	PCT-US93-07347-2	Sequence 2, Appl1
45	80	3.3	399	1	US-08-096-623A-20	Sequence 20, Appl

## ALIGNMENTS

RESULT 1  
US-08-466-583-2  
Sequence 2, Application US/08466583  
Patent No. 5919998  
GENERAL INFORMATION:  
APPLICANT: Bandurski, Robert S.  
APPLICANT: Szeszen, Jędrzej B.  
APPLICANT: Szczygłowski, Krzysztof  
TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
TITLE OF INVENTION: and Plant Growth.  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Greenlee & Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: United States of America  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,583  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/265,427  
FILING DATE: 24-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 11-94A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
TELEX: 49617824  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 471 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-466-583-2

Query Match 31.2%; Score 751.5; DB 2; Length 471;  
Best Local Similarity 38.1%; Pred. No. 1.1e-69;  
Matches 182; Conservative 75; Mismatches 180; Indels 41; Gaps 13;



```

? FILING DATE: 13-AUG-1997
?
? ATTORNEY/AGENT INFORMATION:
?
? NAME: Spector, Eric S.
? REGISTRATION NUMBER: 22495
? TELECOMMUNICATION INFORMATION:
?
? TELEPHONE: 703-415-1500
?
? TELEFAX: 703-415-1508
?
? INFORMATION FOR SEQ ID NO: 2:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 471 amino acids
?
? TYPE: amino acid
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: protein
?
US-09-106-464-2

```

```

RESULT      5
US-08-466-583-5
: Sequence 5, Application US/08466583
: Patent No. 591998
: GENERAL INFORMATION:
: APPLICANT: Bandurski, Robert S.
: APPLICANT: Szerzen, Jędrzej B.
: APPLICANT: Szczygłowski, Krzysztof
: TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
: AND PLANT GROWTH.
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee & Winner, P.C.
: STREET: 5370 Manhattan Circle, Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: United States of America
: ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,583
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/765,427
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Felder, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 11-94A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO.: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 131 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
FRAGMENT TYPE: Internal
US-08-466-583-5

Query Match          16.2%; Score 390; DB 2; Length 131;
Best Local Similarity 55.0%; Pred. NO. 5.4e-33;
Matches 72; Conservative 17; Mismatches 40; Indels 2; Gaps 1.

OY      264 CLEMSTNPRSSVVVVSQSFVNTRKSKQMEELARGILDCGREFLWVRNBESEVYLSCM 323
        | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1 CKRWLDTPDRSVAAYVSPGISLASLGNACQEEELARGLLAGKPFLWVRASDSHQVRYLL 60

OY      324 BELKRVG--KIYSMSCQLFEVLTHPSLGCFVTHCGNNSTLESISFGVPNVAFPMQFDGTN 381
        |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db       61 AEATVTGAAMVYPKCPOLDVLAHPAVGCFTYHICGNNSTLEALSFGVPYAAALMTDOPIN 120

OY      382 AKLMEDVMRTG 392
        | : | | |
Db      121 ARNELAWGAG 131

RESULT      6
PCT-US95-07820-5
: Sequence 5, Application PC/TUS9507820
: GENERAL INFORMATION:
: APPLICANT: Bandurski, Robert S.
: APPLICANT: Szerzen, Jędrzej B.
```

[illegible]











Fri Aug 4 15:19:50 2000

us-09-147-955-6.raii

Page 9

QY	421	EKSrk	425
		:	
Db	431	DKSYK	435

Search completed: August 1, 2000, 19:51:33  
Job time: 18755 sec

[illegible]

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 14:38:19 ; Search time 3616.02 Seconds  
(without alignments)  
2514.661 Million cell updates/sec

Title: US-09-147-955-5  
Perfect score: 2062  
Sequence: 1 attttcccaaaaaataaaaa.....taaaaaaaaaaaaaaaaaa 2062

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*

44: gb\_est25: \*  
45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: gb\_est33: \*  
53: gb\_est34: \*  
54: gb\_est35: \*  
55: gb\_est36: \*  
56: gb\_est37: \*  
57: gb\_est38: \*  
58: gb\_est39: \*  
59: gb\_est40: \*  
60: gb\_est41: \*  
61: gb\_est42: \*  
62: gb\_est43: \*  
63: gb\_est44: \*  
64: gb\_est45: \*  
65: gb\_est46: \*  
66: gb\_est47: \*  
67: gb\_est48: \*  
68: gb\_est49: \*  
69: gb\_est50: \*  
70: gb\_est51: \*  
71: gb\_est52: \*  
72: gb\_est53: \*  
73: gb\_est54: \*  
74: gb\_est55: \*  
75: gb\_est56: \*  
76: gb\_est57: \*  
77: gb\_est58: \*  
78: gb\_est59: \*  
79: gb\_est60: \*  
80: gb\_est61: \*  
81: gb\_est62: \*  
82: gb\_est63: \*  
83: gb\_est64: \*  
84: gb\_est65: \*  
85: gb\_est66: \*  
86: gb\_est67: \*  
87: gb\_est68: \*  
88: gb\_est69: \*  
89: gb\_est70: \*  
90: gb\_est71: \*  
91: gb\_est72: \*  
92: gb\_est73: \*  
93: gb\_est74: \*  
94: gb\_est75: \*  
95: gb\_est76: \*  
96: gb\_est77: \*  
97: gb\_est78: \*  
98: gb\_est79: \*  
99: gb\_est80: \*  
100: gb\_est81: \*  
101: gb\_est82: \*  
102: gb\_est83: \*  
103: gb\_est84: \*  
104: gb\_est85: \*  
105: gb\_est86: \*  
106: gb\_est87: \*  
107: gb\_est88: \*  
108: gb\_est89: \*  
109: gb\_est90: \*  
110: gb\_est91: \*  
111: gb\_est92: \*  
112: gb\_est93: \*  
113: gb\_est94: \*  
114: gb\_est95: \*  
115: gb\_est96: \*  
116: gb\_est97: \*



Oy 1014 gaaaattatcttctgattgaaatgctgacgcatccctggttggagatgt 1073  
 ||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 471 GGAATACTACCATGAGTTCACACTTGAAAGTCTCTGACACATCCATTGTAGGATGTT 530  
 Oy 1074 tctgtacacactcgggtggaattcgtactctagagatatcttctggttccatgtg 1133  
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 531 TTGCTTCGACACTGTGATGATTCGACTCTAGAAAGCCTATCGTACGTACCTGTGG 590  
 Oy 1134 tggcttctcgacgttggcttcgatacgaaggacgaatcgagatgtgagatgtgt 1190  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 591 TGGCATTTCTCATTTGACAGATTCAGGACATTAATGCAAGTTGATGAAGATGTTT 647

**RESULT 2**  
**AM349414/c**  
**LOCUS** AM349414 747 bp mRNA EST 01-FEB-2000  
**DEFINITION** GM210007A20D2R Gm-r1021 Glycine max cDNA 3', mRNA sequence.  
**ACCESSION** AM349414  
**VERSION** AM349414.1 GI:6847124  
**KEYWORDS** EST.  
**SOURCE** soybean.  
**ORGANISM** Glycine max  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;  
 Fabaceae; Papilionoideae; Glycine.  
 1 (bases 1 to 747)  
 Vodka, L., Kelm, P., Shoemaker, R., Retzel, E., Khanna, A., Corryell, V.,  
 Erpelting, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.  
 A Functional Genomics Program for Soybean (NSF 9872565)  
 Unpublished (1999)  
 On Jul 9, 1999 this sequence version replaced gi:5434916.  
 Contact: Vodka, L.O., PI, A Functional Genomics Program for  
 Soybean (NSF 9872565)  
 Lewin, H. A., Director, Keck Center for Comparative and Functional  
 Genomics  
 University of Illinois  
 Edlin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA  
 Tel: (217) 244-6147  
 Fax: (217) 333-4582  
 Email: l-vodka@uiuc.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134. For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or Info@genome  
 systems.com web site: www.genomesystems.com  
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

**FEATURES**  
**source**  
 1..747  
 Location/Qualifiers  
 /organism="Glycine max"  
 /cultivar="Williams"  
 /db\_xref="taxon:3847"  
 /clone\_lib="Gm-r1021"  
 /tissue\_type="root"  
 /lab\_host="X110-Gold"  
 /note="vector: pbluescript II XR; Site\_1: EcoRI. Site\_2:  
 XhoI. Library Gm-r1021 is a sequence-driven, retracted set  
 of the original library Gm-cl004 which was prepared from  
 root cDNA. The mRNA was isolated from entire roots of 8  
 day old 'Williams' seedlings which were propagated on  
 paper towels with distilled water. Stratagene's cDNA  
 Synthesis Kit (catalog #200401) was used to synthesize  
 the cDNA. The Gm-cl004 library was constructed by Dr.  
 Paul Keim & Virginia H. Corryell, Department of Biology,  
 Box5640, Northern Arizona University, Flagstaff, AZ  
 86011, email: paul.keim@uau.edu, virginia.corryell@uau.edu.  
 The contig analysis to select unique genes was performed  
 by the laboratory of Ernest Retzel, Computational Biology  
 Centers, University of Minnesota,  
 http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html  
 . Reracking was performed by Genome Systems, St. Louis,  
 http://www.genomesystems.com, and sequencing by the Keck  
 Center for Comparative and Functional Genomics,

University of Illinois,  
 http://www.life.uiuc.edu/biotech/keck.html."

**BASE COUNT** 206 a 235 c 97 g 187 t 22 others  
**ORIGIN**

Query Match 12.0%; Score 248.2; DB 71; Length 747;  
 Best Local Similarity 62.7%; Pred. No. 2.4e-45;  
 Matches 427; Conservative 0; Mismatches 233; Indels 21; Gaps 3;

Oy 728 attcctccgacatctcttgacggtlaaagaatccctcgacaggtctctcgagagatttg 787  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 733 ATTCNNNNNCGTTCNNNGGTAAGATCCCGCGGATACCTNNNATGCGGTGATNNN 674  
 Oy 788 ttcgagaaagggtcgaatgaacgagatgcctcgaaatggttgaacgaatccctgact 847  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 673 NNC-----NNTGCTCAATGATGTACGTGATGATGTTGGACACAGCCTGATTA 623  
 Oy 848 tgggtgtttagtcttcgttcggaagcttcgttaatacagacgaagtcgaatggaag 907  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 622 TCTGTGTTTATGTTTCATTTGATACCTTGTCTGTCTGTGCTGATATGACAGATGAAGCA 563  
 Oy 908 atagcaagaaggctgttagatttgaggagccgtttcttggttggttgaagtaaacga 967  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 562 CTTGACGCGCGCGTGTGATTCGGATATCTCTTCTTGCGGTCTATTAGATATGCA 503  
 Oy 968 ggaagaaggtattgataatgtcagatgagagagattgaacgaatggggaatgatact 1027  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 502 GGATATGAGAG-----ATPACTGACAGAGAGAACTGACAGAGGCGTAAATGTGMAA 449  
 Oy 1028 tgggttctcaatttgaatccctcgacgcatccctcggttggttggttcttgacacatcg 1087  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 448 TGGTGTCTCAGGTGAGGTTCTGTGCAATGTTCTTGTGGTTGTTTGTATGCAATTC 389  
 Oy 1088 ggttggaattcgactcttagagatatactcttcgggttcgagatgttggttcgcag 1147  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 388 GGTGGAATTCGATTCATGGAAGTTGGGTGCGGGGTTCTATGATGCGGTTCCGCA 329  
 Oy 1148 tggcttcgataaaggagacgaatgcgaagctgtatgagagatgtgtgagacgggtgtgaga 1207  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 328 TGGACACAGCAACGAGCAATGCAAGATGTCGAAGATGTGGAAAGCGGGGTGAGG 269  
 Oy 1208 ggt-----agagctatgaaggaggttaagcgtgtgtgattgagatgaatgaatt 1261  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 268 GTGATGATTAAGTGAATGTGAGAGAGATGTTGAAGCAAGAGATTAAGAAATG 209  
 Oy 1262 attgaagagttatgaattgaggagaaagatgaagaaacttagagagaggtctgcaag 1321  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 208 TTGATGCTGTCTCATGGGAGTGGAGAAAGACAGAAATTCAGAGATCTGATTA 149  
 Oy 1322 tggaaagatttgcaagaagaactatgaggaagatgatacttaacaacctcaag 1381  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 148 TGGAAATGCTTGGCCAGGGAAGCCGTCACGGAAGAGAGCTCTCGATGTAATATGAG 89  
 Oy 1382 gttcttctgattgatttga 1402  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 88 ACTTTCTCCATGATGTGTCA 68

**RESULT 3**  
**A1771830**  
**LOCUS** A1771830 536 bp mRNA EST 29-JUN-1999  
**DEFINITION** EST529330 tomato ovary, TAUO Lycopersicon esculentum cDNA clone  
 CLEB38617, mRNA sequence.  
**ACCESSION** A1771830  
**VERSION** A1771830.1 GI:5269871  
**KEYWORDS** EST.  
**SOURCE** tomato.  
**ORGANISM** Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;  
 Solanaceae; Solanum; Lycopersicon.

**REFERENCE**  
 1 (bases 1 to 536)



```

OY 840 ctgcattcttggttggttaacgttctgctggaagcttctgtatacgaagagtcgcaaa 899
    || || || || || || || || || || || || || || || || || || || || ||
DB 284 CTAATATCATCAATTCCTATATATCATTTGGAGTCTATTAATTAATCAACAAACCAA 343
OY 900 tggagaagatacgaagaagcgtctagatttgcggagagccgttttctgtggttgtaagag 959
    || || || || || || || || || || || || || || || || || || || || ||
DB 344 AGGAGAGATATTCGAAGAGGTTGATAGACATCCAAAGGCCATCTTATGGGTAAATPAGAG 403
OY 960 taaacgaag-----gagaagaaggtattgataagttgcattggaaggagtgaacagatgag 1013
    || || || || || || || || || || || || || || || || || || || || ||
DB 404 ATCAGAGAAGCAAGAAAGAGAAAGAAATTCAGTTGCATGATGATGATGAGAAAGCAAG 463
OY 1014 ggaataatgtatctgtggttctcaattggaagtcctgacagatccctctgtggatgct 1073
    || || || || || || || || || || || || || || || || || || || || ||
DB 464 GGAATAATGATCCATGCTGTTCACAACTTGAAAGTCTGACATCCATCTTTAAGATGTC 523
OY 1074 tctgacacacactgcgggtggaat 1096
    || || || || || || || || || || || || || || || || || || || || ||
DB 524 TTGTCTCGCAGCTGTGTGATGAT 546

RESULT 5
LOCUS A1899102 478 bp mRNA EST 27-JUL-1999
DEFINITION EST268545 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
ACCESSION A1899102
VERSION A1899102.1 GI:5605004
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 478)
AUTHORS Holt,I.E., Liang,F., Upton,J., Romling,C.M., Craven,M.B.,
Fuji,I.C.Y., Bowman,C.L., Niernan,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
On Feb 18, 1999 this sequence version replaced gi:4296853.
COMMENT JOURNAL
Contact: David Frisch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfri@clcmson.edu
5 prime sequence.
FEATURES
source
1..478
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="GLED37A16"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI, Site_2:
XhoI; cled - Tomato Carpel EST library. OligodT-primed and
directionally cloned cDNA in vector lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."

BASE COUNT 172 a 67 c 102 g 137 t
ORIGIN
Query Match 9.9%; Score 203.2; DB 45; Length 478;
Best Local Similarity 66.9%; Pred. No. 2,4e-35;
Matches 324; Conservative 0; Mismatches 148; Indels 12; Gaps 2;

```

```

OY 627 gtgaagaanaaacctaaggtcttggtgaacagctttagatcgttggagcctatgcctca 686
    || || || || || || || || || || || || || || || || || || || || ||
DB 1 GTGAAGAAATCCAAAGGACTTGTGAAATATCATTTGATCATTTAGAGCCCTCA 60
OY 687 agggcattataagatcagagatgcatgcatcgccgctttagcttccttcagattcttg 746
    || || || || || || || || || || || || || || || || || || || || ||
DB 61 AAGCTATTTAAATAATACAAATTTAATTGCAATTTGACCATTTGATCTTATCATCTTTGG 120
OY 747 acgtaaatgatcccttcgacagagcttcgcggagattgttcgaaagggtcgaatg 806
    || || || || || || || || || || || || || || || || || || || || ||
DB 121 GTGGAATAATCATTTGAAATCTTCAATTTGTGTGATCATCTTTTTCAAA-----CTCAA 174
OY 807 agcagatctgcgtaattgttgtagcagaaacccctgcattctcgtgtgtttagcttctg 866
    || || || || || || || || || || || || || || || || || || || || ||
DB 175 ATGATGACTACATGAAATGATTAACCAACCAACCAACCAATGTTATATCTCAT 234
OY 867 tgggaagcttctgtaatacgaagaagtcgcaaatggaagaagatagcaagaagcgttag 926
    || || || || || || || || || || || || || || || || || || || || ||
DB 235 TTGGGAGTCTATTTGAATTTATCAAGAAACCAAGAGAGATGCAAAAGGTTGATAG 294
OY 927 attgtggagagccgttttctgtggttgtaag-----agtaacgaagaagaagagtat 980
    || || || || || || || || || || || || || || || || || || || || ||
DB 295 AGATCCAAAGGCCATTTCTTATGCGTAAATAAGATCAAGAGAGAGAAAGAAAGAGAGA 354
OY 981 tgataagttgcattggaagagttgaaacgagttggaagaatttattcttcttcaat 1040
    || || || || || || || || || || || || || || || || || || || || ||
DB 355 AATTAGTTGCATGATGATTAAGCAAGCAAGGAAATATGATCATCTGTTCACAAAC 414
OY 1041 tggaaagctgcagacatccctctgtggaattcttctgacacactgcgggtggaattcga 1100
    || || || || || || || || || || || || || || || || || || || || ||
DB 415 TTGAAGTCTTGACACATCATCTTTAAGATGTTGTCTCGCAGCTGTGATGGAATTCGA 474
OY 1101 ctct 1104
    || || || || || || || || || || || || || || || || || || || || ||
DB 475 CTTCT 478

RESULT 6
LOCUS A1896013 559 bp mRNA EST 27-JUL-1999
DEFINITION EST265456 tomato callus, TAMU Lycopersicon esculentum cDNA clone
ACCESSION A1896013
VERSION A1896013.1 GI:5601915
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 559)
AUTHORS Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S.,
Romling,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfri@clcmson.edu
5 prime sequence.
FEATURES
source
1..559
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="GLED37A16"
/clone_lib="tomato callus, TAMU"

```





```

Db      536 ATATCATTTGGAGTCTATTGAATTATCAAGAAACCAAGAGGATTCGAAGAAG 595
Qy      918 ggcgttagatgttgaggccgcttttctgtggtgtaagagtaacgaagag 971
Db      596 GGTGTATGAGATGCCAAAGCCATCTTATGGTATATAGATCAAGAAGAAG 649

RESULT  8
LOCUS   A1487571 520 bp mRNA EST 29-JUN-1999
DEFINITION EST245893 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
ACCESSION A1487571
VERSION  A1487571.1 GI:4382942
KEYWORDS EST.
SOURCE  tomato.
ORGANISM Lycopersicon esculentum
          Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
          Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
          Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 520)
AUTHORS  Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
          Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
          Fujii,C.Y., Bowman,C.L., Niernan,W., Fraser,C.M., Venter,J.C.,
          Martin,G.B., Tanksley,S.D. and Giovannoni,J.
          Generation of ESTs from tomato carpel tissue
          Unpublished (1999)
TITLE     On Apr 7, 1998 this sequence version replaced g1:3035667.
JOURNAL   Contact: David Frisch
          Clemson University Genomics Institute
          100 Jordan Hall, Clemson, SC 29634, USA
          Tel: 864 656 4366
          Fax: 864 656 4293
          Email: dfri@che.clemson.edu.
FEATURES
          source 1..520
                  Location/Qualifiers
                  /organism="Lycopersicon esculentum"
                  /cultivar="TA496"
                  /db_xref="taxon:4081"
                  /clone="cLED10D21"
                  /clone_1lb="tomato ovary, TAMU"
                  /tissue_type="carpel"
                  /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                  /lab_host="XLI-Blue MRF"
                  /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
                  XhoI; cLED - Tomato carpel EST library. OligodT-primed and
                  directionally cloned cDNA in vector lambda Zap II with 5'
                  and 3' ends located at the EcoRI and XhoI sites,
                  respectively."

BASE COUNT 194 a 76 c 107 g 143 t
ORIGIN
Query Match 9.2%; Score 189.8; DB 40; Length 520;
Best Local Similarity 66.1%; Pred. No. 2,3e-32;
Matches 310; Conservative 0; Mismatches 147; Indels 12; Gaps 2;

```

```

Db      237 GGTATCTTTTCAAAA-----GTCAATGATGACTACATGCAATGTTAAACCAAGC 290
Qy      840 ctgcattctcggtgttacctgttcggaagcttcgtaacgaagtcgcaaa 899
Db      291 CTAAATCATCAATGTGTTATATCTCATTTGGGAGTCTATTGAATTATCAAGAACCAAA 350
Qy      900 tggaaagatagcaagaaggcctgttagattgtgaggagccgcttttgggtgtgtaag-- 957
Db      351 AGGAGGAGATGCAAAAAGGTTGATAGATCCAAAGGCCATCTTATGGTATATAGAG 410
Qy      958 ----agtaacgaaggagaagagttatgaagcttcgatggaagttgaacgagttg 1013
Db      411 ATCAAGACAGAAAAAGAAAGAAAGAAATTAAGTTGCAATGCAATTAAGAAACCAAG 470
Qy      1014 ggaanaattglatcttgggttctcaattggaagtcctgacatccctc 1062
Db      471 GGAATAATGATACCATGTGTTCACAACTTGAAGTCTGACATCATC 519

RESULT  9
LOCUS   AM034683 685 bp mRNA EST 15-SEP-1999
DEFINITION EST278414 tomato callus, TAMU Lycopersicon esculentum cDNA clone
ACCESSION AM034683
VERSION  AM034683.1 GI:5893439
KEYWORDS EST.
SOURCE  tomato.
ORGANISM Lycopersicon esculentum
          Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
          Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
          Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 685)
AUTHORS  Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
          Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S.,
          Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
          Giovannoni,J.
          Generation of ESTs from tomato callus tissue
          Unpublished (1999)
TITLE     On Apr 3, 1998 this sequence version replaced g1:3018868.
JOURNAL   Contact: David Frisch
          Clemson University Genomics Institute
          100 Jordan Hall, Clemson, SC 29634, USA
          Tel: 864 656 4366
          Fax: 864 656 4293
          Email: dfri@che.clemson.edu
FEATURES
          source 1..685
                  Location/Qualifiers
                  /organism="Lycopersicon esculentum"
                  /cultivar="TA496"
                  /db_xref="taxon:4081"
                  /clone="cLEC11D5"
                  /clone_1lb="tomato callus, TAMU"
                  /tissue_type="callus"
                  /dev_stage="25-40 days old"
                  /lab_host="XLI-Blue MRF"
                  /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
                  XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
                  of seedlings 7-10 days post-germination were excised, cut
                  at both ends and placed on MS medium with no selection.
                  Mixed callus was harvested at 25 and 40 days and included
                  undifferentiated masses. Tomato Callus EST library"

BASE COUNT 219 a 128 c 137 g 201 t
ORIGIN
Query Match 8.9%; Score 183.4; DB 63; Length 685;
Best Local Similarity 59.9%; Pred. No. 6,4e-31;
Matches 379; Conservative 0; Mismatches 221; Indels 33; Gaps 3;

```

```

QY 345 tcaagctctggtgactccaccctcttgatcgtgagcgccaaagtgcgctgagttcc 404
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 59 TCAGCTCCCTGCTCTACACTCTTCTGCTACTTGGGAGCTGAGTAGTACGGCGTGAACCTCC 118
QY 405 atcccgagagcgagctactctggtatgagccagctaaagctggtgataatctactttc 464
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 ATATCCCATCGCGCTTACTATGATTCACACGACAGCTGCTGACATATACATCTATTT 178
QY 465 attcacaagctatagtagcagcaaat-----cagtcggtgtcggatgctatc 512
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 ACTTCATGCTGCTATGAGATGATAAATGAAGTACTAGCTCAATATGATCCAAATGGAGTA 238
QY 513 acttgcggagagagctccagctgctgcccagcgtagattacgcgtctctctctcctt 572
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 TCCAATTTGCCAAGCGCTTCATTTACTTAAAAAGCCAAGATCTTCATCTTTTATTAGTTTCAT 298
QY 573 c-----cagcagtagagatcgcgttcaactgtagaaggagaattggaaa 617
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 299 CTAGCTCAAAAGATGATTAAGTATAGTTTCTCTACCAACATTCMAAGACACTAGACA 358
QY 618 cttaagaaggtgaagaaacctaaagctctggtgaacagcttgatgagcttgagagcgtg 677
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 CATTTAGTGGTGAAGAAATCCAAAGCTACTTGTGAAATACATTGATGATTAAGACTAG 418
QY 678 atgcgtcacaagggcactgataagtagcagagatgtagcgtgagcgcttgatctctcgcg 737
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 419 AGGCACCTCAAGCTATTGAAATAATCAATTTAATTGGAATGGACATTTGATCTCTTCAT 478
QY 738 catctctgagagcgtaaaagatcctctcgacagagctcttcggcgagagatctgtctcgagaaag 797
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 479 CATTTCTGGGCGGAAGATTCATTTGGAATCTGGAATCTTTGGTGTATCTTTTAAAAA-- 536
QY 798 gttcgaatgtagcagatctgctcgatctggtgagcagcaatccctcgatctcggtgttt 857
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 537 -----GTCAATGATGACTACTACTGAAATGCTTAACACAAACCTTAATCAATTTGTTT 592
QY 858 acgttctgctcgaagctctgtaatacgcagaaagctgcaaatgtagaagataagcaag 917
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 593 ATATCTCATTTGGGAGCTCTATTGAAATTTATCAAGAAACCAAGAGAGATTTGCAAAAAG 652
QY 918 ggtcgttagatctgtggagagcgcttctgtggg 950
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 653 GGTGATAGAGATCCAAAGCCATCTTATGCG 685

```

RESULT 10

LOCUS A1489014 500 bp mRNA EST 29-JUN-1999

DEFINITION EST247353 tomato ovary, TAMU Lycopersicon esculentum cDNA clone

ACCESSION A1489014

VERSION A1489014.1 GI:4384385

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 500)

Auclair, J., Vredelov, J., White, R., Matern, A.L., Vison, T., Holt, I.E., Liang, F., Upton, J., Romling, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and Giovannoni, J. Generation of ESTs from tomato carpel tissue

Unpublished (1999)

On Nov 2, 1998 this sequence version replaced gi:3830982.

Contact: David Frisch

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 4366

Fax: 864 656 4293

Email: dfrisch@CLEMSON.EDU.

```

FEATURES
    source          Location/Qualifiers
                    1..500
                    /organism="Lycopersicon esculentum"
                    /cultivar="T496"
                    /db_xref="taxon:4081"
                    /clone="CLEDI9F7"
                    /clone_1id="tomato ovary, TAMU"
                    /tissue_type="carpel"
                    /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                    /lab_host="X11-Blue MR"
                    /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector lambda Zap II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."
BASE COUNT      188 a      69 c      105 g      138 t
ORIGIN
Query Match      8.4%; Score 174; DB 40; Length 500;
Best Local Similarity 65.1%; Pred. No. 7,7e-29;
Matches 293; Conservative 0; Mismatches 145; Indels 12; Gaps 2;

```

QY 600 tgaagagaatctgaaacttgaagctgaaagaaacccaagctctgtgtagcagct 659

Db 57 TCAAGAGACACATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 116

QY 660 ttgatcgctgagagcgatgagctgagcagcattgataagtagagatgtagcagctg 719

Db 117 TTGATGATTTAGACCTAGAGCCACTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAG 176

QY 720 ggcgtgactcctctccgactctctgagcgtgaaagatcctcggacagcgtcttcgagc 779

Db 177 GACCATTCATTCCTTCATCATCTTTGGTGGGAAAGATTCATTTGGAATCTTCATTTGGTG 236

QY 780 gagatttctgagaagaggtcgaaatgtagcagatctgctcgaatggttgcagcaatc 839

Db 237 GTGATCTTTTCAAAA-----GTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 290

QY 840 ctgacatctcggtgttaagcttctgctgagagcttctgtaatacgcagagctgcgaa 899

Db 291 CTAAATCATCATTTGTTATATCTCATTTGGAGCTATTTAATTTATCAAGAAACCAAA 350

QY 900 tgaagagatagcaagagcgctgttagatctgtgagagcgcttctgtggtgtgaaag 957

Db 351 AGGAGAGATTCGCAAAAGGCTGATAGATCCAAAGCCATTCTTATGGTATATAGAG 410

QY 958 ----agtaacgaaggaagaggtatgataagctgtagagaggtgtagcagagctg 1013

Db 411 ATCAAG 470

QY 1014 ggaagaatgtatctctgtgtgtctcaatctg 1043

Db 471 GGAAGATATGATCATGCTGTTCAACAATTTG 500

RESULT 11

LOCUS AW651280 658 bp mRNA EST 04-APR-2000

DEFINITION EST329734 tomato germinating seedlings, TAMU Lycopersicon

ACCESSION AW651280

VERSION AW651280.1 GI:7412518

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 658)

Auclair, J., Vredelov, J., White, R., van der Hoeven, R.S., Holt, I.E., Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Romling, C.M., Nierman, W., Fraser, C.M., Giovannoni, J.J., Martin, G.B. and

Tanksley, S.D.  
Generation of ESTs from germinating tomato seed  
Unpublished (2000)  
On Jan 6, 2000 this sequence version replaced gl:6677182.  
COMMENT

Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU

FEATURES  
source

Location/Qualifiers  
1..658  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLE116K15"  
/clone\_1ib="tomato germinating seedlings, TAMU"  
/tissue\_type="whole seedlings"  
/dev\_stage="7 days post imbibition"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; 7 days post imbibition on water agar. Mixed stage whole germinating seedlings from seed coat emergence up to two centimeters in seeds not showing obvious signs of germination were discarded."

BASE COUNT 208 a 126 c 124 g 200 t  
ORIGIN

Query Match 7.8%; Score 161.6; DB 79; Length 658;  
Best Local Similarity 58.0%; Pred. No. 4.6e-26;  
Matches 385; Conservative 0; Mismatches 239; Indels 40; Gaps 4;

QY 248 gatggaagaactacatgctgcagatgaataaagcgttagcgatcact 307  
DB 2 GATTCTAAACGTACACGTACAGAGAAAGTCGTGGCCCAATGAGGATGTC 61  
QY 308 ctggcagcaataatgctgatacaaaaagcaaaatcagctgtggtactccac 367  
DB 62 ATTTTGAAGGTCACAGC-----AAGGACGCGCTGTCAGTCCTCGTACACTCTT 115  
QY 368 ctcttgatggcgccgaagctgctgagctgcatcccgagcgagctactcg 427  
DB 116 CTGCTACCTTTGGGACAGTGTAGCGCGTGAATCCATATCCATCCGCTACTATGA 175  
QY 428 attgacccagctacgctgtgtgatatacttatttcaacgagctataagcagaa 487  
DB 176 ATTCACACGACCACTGTGTAGACATATCTACTTACTTCAATGGCTATGAGATGAA 235  
QY 488 at-----cgatggggttcgagctatcacttgcgcggaagatcccgatg 535  
DB 236 ATGAAGTGTACTACGTCMAATGATCCAAATGGAGTATCCAAATGGCCAAAGCTTCCATT 295  
QY 536 ctggccagcgatgattacgcgtctctctctccctccagcatgagagatccgltca 595  
DB 296 CTAAAGGCCAATCTCTTCATCTTTTATGTTTCATTCATGCTCAAAATATGATGATA 355  
QY 596 ctg-----atgaagagagaattgaaacttgaaggtgaagaaacc 639  
DB 356 TACTTTTGCTCTACCAACATTCMAAGACAACTAGACACATTTAGATGTTGAAGAAAAATCC 415  
QY 640 taagctctggggaagaagcttgcgtgtgagcctgagcgtcaagcacttgataa 699  
DB 416 AAAGGTAATGTAATACATTTGATGATGATGATGATGATGATGATGATGATGATGATG 475  
QY 700 gtaagagatgataatgacacgagcgtgattctctccgcatcttgaagcgttaagatcc 759  
DB 476 ATTCATTTTAATTTGAATTTGACATTTGATTTCTTCACTATTTTGGGTGGAAGATTC 535  
QY 760 ttgcgaagagctcttcggcgaggaattgttcgagaagaggtcgaatgacgagatgct 819  
DB 536 ATTCGATCTTCAATTTGGTGTGATCTTTTCAAAA-----GTCAATGATGATGATCAT 589

QY 820 cgaatggtgagcagcaatccctgacatcttcggtgttaccgttctcggaagcttcgt 879  
DB 590 GGAATGCTTAACACCAAGCTTAATCATCATATTTATATCTCATTTAGGAGCTATT 649  
QY 880 taat 883  
DB 650 GAAT 653

RESULT 12  
A1483541 476 bp mRNA EST 29-JUN-1999  
LOCUS EST249390 tomato ovary, TAMU Lycopersicon esculentum cDNA clone  
DEFINITION CLE025M13, mRNA sequence.  
ACCESSION A1483541  
VERSION A1483541.1 GI:4387465  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;  
Solanales; Solanum; Lycopersicon.  
1 (bases 1 to 476)  
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T.,  
Holt, I.E., Liang, F., Upton, J., Ronning, C.M., Craven, M.B.,  
Fujii, C.Y., Bowman, C.L., Tanksley, S.D. and Giovannoni, J.  
Martin, G.B., Tanksley, S.D. and Giovannoni, J.  
Generation of ESTs from tomato carpel tissue  
Unpublished (1999)  
Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU.

FEATURES  
source

Location/Qualifiers  
1..476  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLE025M13"  
/clone\_1ib="tomato ovary, TAMU"  
/tissue\_type="carpel"  
/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"  
/lab\_host="X11-Blue MRP"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."

BASE COUNT 180 a 64 c 101 g 131 t  
ORIGIN

Query Match 7.7%; Score 159.6; DB 40; Length 476;  
Best Local Similarity 65.8%; Pred. No. 1.2e-25;  
Matches 250; Conservative 0; Mismatches 124; Indels 6; Gaps 1;

QY 600 tgaaggaagaattggaacttgaaggtgaagaaacttaaggtcttgtagaagcgt 659  
DB 57 TCAGAGGCAACTTAGACACTTTAGATGTTAGAGAAAATCCAAAGGTACTTGCAATCAT 116  
QY 660 ttgagctgttgagcctgagcgtcgaagccatgataagctacgagatgattgcaatcg 719  
DB 117 TTGATGATTAAGAGCTAGACCACTCAAGCTATTGAAAATATCAATTTAATTTGAATG 176  
QY 720 ggcgctgattctctccgcatcttggagcgttaagatccttggagaagcttcctggcg 779  
DB 177 GACCATTTGATTCCTTCATCTTTGTTGGTGAAGAAATTCATTTGATTTGATTTGCTG 236  
QY 780 gagattgttcggaagaggtcgaatgagaagcattgcccgaatggttgagcagcagatc 839

Db	Accession	Score	Length	Gap
Db	237 GCGATCTTTTCAAAA-----GTCAAATGAGTACTACTGGAATGCTTAAACCAAGC	290		
Qy	840 ctcgaatcttcggtggttaacgttcgttcggaagcttcgtaatacagcgaagtcgcaaa	899		
Db	291 CTAATCATCAATTTGTTTATCATCTATTGGAGTCTATTTGAATTTATCAAAACAAACAA	350		
Qy	900 tggaaagagatgcaagaagagcgctgttagattgtgggaagcgcttttgggtggtgaag	959		
Db	351 AGGAGAGATGTCGCAAAAGGGTGAATGAGATCCAAAGCCATTCTTATGGAATTTAGAG	410		
Qy	960 taaacgaagagaagagga 979			
Db	411 ATCAAGACAGCAAAAGCA 430			
RESULT	13			
LOCUS	A1779099	517 bp	mRNA	EST
DEFINITION	EST259978 tomato susceptible, Cornell Lycopersicon esculentum CDNA			
ACCESSION	A1779099			
VERSION	A1779099.1	GI:5277140		
KEYWORDS	EST.			
SOURCE	tomato.			
ORGANISM	Lycopersicon esculentum			
REFERENCE	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; eunasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.			
AUTHORS	D'Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E., Liang, F., Updon, J., Romling, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nlerman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D., Giovannoni, J.J. and Martin, G.B.			
TITLE	Generation of ESTs from Pseudomonas susceptible tomato			
JOURNAL	Unpublished (1999)			
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2285531.			
FEATURES	<p>Contact: David Fritsch Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfritsch@CLEMSON.EDU</p> <p>5 prime sequence.</p> <p>Location/Qualifiers</p> <p>1..517</p> <p>/organism="Lycopersicon esculentum"</p> <p>/cultivar="R11-13 (Rio Grande x Money Maker)"</p> <p>/db_xref="taxon:4081"</p> <p>/clone="CLES719"</p> <p>/clone_11b="tomato susceptible, Cornell"</p> <p>/tissue_type="leaf"</p> <p>/dev_stage="4-week old"</p> <p>/lab_host="SOLR"</p> <p>/note="vector: pbluescript SK(-); Site_1: EcoRI, Site_2: XhoI; cLES - tomato Pseudomonas susceptible EST library. directionally cloned cDNAs inserted into pbluescript SK(-) at 5' end with EcoRI and 3' end with XhoI site"</p>			
BASE COUNT	185 a 78 c 102 g 152 t			
ORIGIN				
Query Match	7.7%; Score 158; DB 44; Length 517;			
Best Local Similarity	66.1%; Pred. No. 2, Be-25;			
Matches	246; Conservative 0; Mismatches 120; Indels 6; Gaps 1;			
Qy	600 tgaagagagaatctggaacttgaagagtggaagaaacctaagcttggtgaacagct 659			
Db	152 TCAAGAGCAATAGACATATGATGCTGAGAGAAATTCAAAGGTACTTGGATATCAT 211			
Qy	660 ttgatagcgtttgagcctgctgctcgaagccattgataagtagagatgattgcacatg 719			

Db	212	TTGATGCAATTAGAGCTAGAGCCACGTCACAAAGCATTTGAAAAATACATTTTAATGGAATTG	271
Oy	720	ggcggctgattccctcgcgcatctcttgagcgggtaaagatccctcggacaggtcttcggcg	779
Db	272	GACCAATGATTCCTTCATCATCTTGGGGGTGGAAAAAGATTTCATTTGGAATCTTCATTTGGGTG	331
Oy	780	ggagattcttcgagaaagggttcgaattgcagcagcatctgcttcgaattggttcgacgaatc	839
Db	332	GGGATCTTTTTCAAAA-----GTCAAATGATGAGCATACATGGAAATGGTTAAACACAAAGC	385
Oy	840	ctcgatcttcggttggttcttaacggttcttcggaagcttcgtaatacgaagtcgacaa	899
Db	386	CTAATTCATCATTTTATTATATCATTTTGGAGGCTATTTGATTTATTCACAAACACAA	445
Oy	900	tggagagatagcagaagggctgcttagattcttgaggagccggttttcttggttgtaagag	959
Db	446	AGGAGAGCATTTGCCAAAGGGTGTATAGAGATTCAAAGGCCATTTTATGGGTAAATAGAG	505
Oy	960	taaacgaaggag	971
Db	506	ATCAAGAAGAGC	517

  

RESULT	14
AA459541	619 bp mRNA EST 24-FEB-2000
LOCUS	sh2h3.0.y1 Gm-cl017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION	Gm-cl017-4398 5' similar to TF:022820 022820 INDOLE-3-ACETATE
ACCESSION	AA459541
VERSION	AA459541.1 GI:7029758
KEYWORDS	EST.
SOURCE	soybean.
ORGANISM	Glycine max
REFERENCE	Eunaryota, Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
AUTHORS	1 (bases 1 to 619) Shoemaker, R., Kelm, P., Vodkin, L., Expelding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Pearson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
TITLE	Public Soybean EST Project
JOURNAL	Unpublished (1999)
COMMENT	On Apr 7, 1998 this sequence version replaced gi:3034755. Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 for further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com Seq primer: -40RP from Glibco High quality sequence stop: 348. Location/Qualifiers 1..619 /organism="Glycine max" /db_xref="taxon:3847" /clone="GENOME SYSTEMS CLONE ID: Gm-cl017-4398" /clone_id="Gm-cl017" /tissue_type="vegetable buds of field grown plants" /lab_host="XL10-Gold" /note="Vector: pluvscript II XR; Site 1: EcoRI; Site 2: XhoI. This cDNA library was constructed from mRNA isolated from vegetable buds of field grown plants. The cDNA library was prepared using the Stratagene pluvscript II



Fri Aug 4 15:19:49 2000

us-09-147-955-5.rst

Page 12

Job time: 62700 sec

---

Result No.	Score	Query Match	Length	DB	ID	Description
1	2043	99.1	2043	7	AB013598	AB013598 Verbena h
2	700.2	34.0	1507	7	AB013596	AB013596 Perilla f
3	680	33.6	1458	7	AB013597	AB013597 Perilla f
4	424.2	20.6	1864	7	AB027455	AB027455 Petunia x
5	310.6	15.1	19894	8	ATCNR1V38	AL161538 Arabidops
6	310.6	15.1	200576	7	ATPCAO	Z97335 Arabidops
7	288.6	14.0	1594	7	AB000623	AB000623 Nicotiana
8	199.4	9.7	195068	8	ATCNR1V42	AL161542 Arabidops
9	199.4	9.7	197419	8	ATCNR1V41	AL161541 Arabidops
10	199.4	9.7	205065	7	ATPC4	Z97339 Arabidops
11	197.8	9.6	1589	8	ATU81293	U81293 Arabidops
12	187	9.1	103223	8	AC007153	AC007153 Arabidops
13	183.6	8.9	99188	9	AC006533	AC006533 Arabidops
14	179.2	8.7	84203	49	AC0005106	AC0005106 Genomic s
15	178.4	8.7	1513	5	A65232	A65232 Sequence 28
16	178.4	8.7	1513	5	A72396	A72396 Sequence 28
17	176.8	8.6	1459	5	A6525	A6525 Sequence 30
18	172.8	8.4	1969	5	A6529	A6529 Sequence 34
19	170.4	8.3	1440	49	AF190634	AF190634 Nicotiana
20	159	7.7	122338	8	AC002396	AC002396 Arabidops
21	157.6	7.6	808138	7	AB019232	AB019232 Arabidops
22	155	7.5	1588	5	A65256	A65256 Sequence 31
23	147	7.1	2043	7	AB013598	AB013598 Verbena h
24	145	7.0	94786	49	AC0002333	AC0002333 Arabidops

25	143	6.9	566	5	A62509	A62509 Sequence 14
26	143	6.9	566	5	A72382	A72382 Sequence 14
27	141	6.8	1416	7	M2E1AGLU	L34847 Zea mays 1A
28	135	6.5	568	5	A62510	A62510 Sequence 15
29	135	6.5	568	5	A72383	A72383 Sequence 15
30	126.8	6.1	107491	49	AC002391	AC002391 Arabidops
31	126.4	6.1	1732	7	AB033758	AB033758 Citrus un
32	115.2	5.6	1585	7	D85186	D85186 Gentiana tr
33	108.6	5.3	1688	7	AB031274	AB031274 Scutellar
34	106.4	5.2	1619	8	NTU32643	U32643 Nicotiana t
35	100.2	4.9	1786	7	AB038248	AB038248 Ipomoea b
36	97.6	4.7	100746	8	AC005489	AC005489 Genomlc s
37	97.4	4.7	1687	7	AB016819	AB016819 Arabidops
38	97.4	4.7	48282	53	AC024228	AC024228 Arabidops
39	96.8	4.7	1624	8	NTU32644	U32644 Nicotiana t
40	96.4	4.7	1624	5	A68001	A68001 Sequence 1
41	96.4	4.7	1624	7	LETW11	X85138 L.esculentu
42	96	4.7	1578	5	E12713	E12713 Solanum mel
43	94.8	4.6	1665	7	SMGT	X77369 S.melonena
44	94.8	4.6	1665	7	AB009370	AB009370 Vigna mun
45	94.2	4.6	81642	49	AC005851	AC005851 Arabidops

## ALIGNMENTS

RESULT 1  
AB013598  
LOCUS 2043 bp mRNA PLN 20-MAR-1999  
DEFINITION Verbena hybrida HcT8 mRNA for UDP-glucose:anthocyanin  
5-O-glucosyltransferase, complete cds.  
ACCESSION AB013598  
VERSION 1 GI:4115562  
KEYWORDS HGT8: UDP-glucose:anthocyanin 5-O-glucosyltransferase.  
SOURCE Verbena hybrida petal cDNA to mRNA.  
ORGANISM Verbena x hybrida  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Asterales; Gentianaceae; Lamiales; Verbenaceae; Verbena.  
1 (sites)

REFERENCE  
AUTHORS Yamazaki, M., Gong, Z., Fukuchi-Mizutani, M., Fukui, Y., Tanaka, Y.,  
Kusumi, T., and Saito, K.  
TITLE Molecular cloning and biochemical characterization of a novel  
anthocyanin 5-O-glucosyltransferase by mRNA differential display  
for plant forms regarding anthocyanin  
J. Biol. Chem. 274 (11), 7405-7411 (1999)  
JOURNAL 99167509  
MEDLINE 2 (bases 1 to 2043)  
REFERENCE Yamazaki, M., Saito, K. and Gong, Z.  
AUTHORS Direct Submission  
TITLE Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Maml  
Yamazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. &  
Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan  
(E-mail: mamlyep.chiba-u.ac.jp, Tel: +81-43-290-2905,  
Fax: +81-43-290-2905)  
JOURNAL Location/Qualifiers  
FEATURES  
source 1..2043  
/organism="Verbena x hybrida"  
/db\_xref="taxon:76714"  
/issue\_type="petal"  
26..1411  
/gene="HGT8"  
26..1411  
/gene="HGT8"  
/codon\_start=1  
/product="UDP-glucose:anthocyanin 5-O-glucosyltransferase"  
/protein\_id="BAA36423.1"  
/db\_xref="GI:4115563"

/translation="MSRAHYLATPFAAGHINPALQFAKRIANADIOVTFTSVYAMR  
RMSRTAAGSNGLINPYSFSDGYDDGAPROBDDGNSSEKMSRIRKLSPTLANVND  
KSKITFVYSHLFAMAKVAFERHLSALWLEPVALIDIFITNGSDELDASD  
AHLPGGLVLAORDLPFLPSTHERFSLMEKLETELEGKPPVLVNSFALPSP  
AKAIDKYEIMAIPLIPSAFLDGDPSRDSFGDLFEKGSNDDELMLSTNPRSSV

QY	1	attttcccaaaaaataaaaaaagacagagcgcagctctctgccaattccc	60	VYVSGSFVNTTKSQMEETARGLLDCGRPLVWVRYNEGEVLIISCEBELKRYGKIVS WCSOLEVLRHPSIGCVTHGCGNMSTLESTISFGVPVAFQWMDQGNALIMDVWRTG VVRANRSGSVYVDGDEIRRCIRIEVYMDGKSRKLRASAKMKMDLARKNAEDGSSVNN LKVFIDEVVGIT*
DB	1	ATTTTACCAAAAAATATAAAAAATAGACAGAGCGTCACGCTCTTGCGCAATCC	60	polyA_site /note="19 a nucleotides" BASE COUNT 589 a 354 c 508 g 592 t ORIGIN
QY	61	agcaagagacacataaatccgcgcccttcaattcgccaagcgtctcgaatgcgaacat	120	Query Match 99.1%; Score 2043; DB 7; Length 2043; Best local Similarity 100.0%; Pred. No. 0; Matches 2043; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB	61	AGCAGAGGACACATAATATCCGCCCTTCAATTCGCCAAGCGTCTGCCAATGCCGACAT	120	
QY	121	tcaagtcacattctcaccagcgtctcagcatgagcgccatgctcgaacccgcgtg	180	
DB	121	TCAAGTCACATTTCTTCACGCGCTCTACGATGCGCGCATGCCAGAACCGCGCTGG	180	
QY	181	ctcaaacggcgtcactcaattgtgtcgtttccgagcggtatagaagcgggttaagcc	240	
DB	181	CTCAAAAGGCGCTCATCAATTTGTGTGCTTTCCGACGGGTATGACACGGTTACAGCC	240	
QY	241	cggagacgatgagaaagacacatgctcagagatgaaagcagaaggtataaaagccttga	300	
DB	241	CGGAGACGATGGGAAGAACTACATGTCGAGATGAAAAAGCAGAGCTATAAAGCCTTGAG	300	
QY	301	cgatctcttcgagcccaataatgctcaaaaaaagcagaacaaatcagctgtgtgtgta	360	
DB	301	CGATCTCTTTCGACGCCATATATGTCGATCAAAAAAGCAGCAAACTCGTTCGTGTGTA	360	
QY	361	ctccaccctcttcagtcagggcgagcgaagtcgagcgtgagttccatctccggaagcggt	420	
DB	361	CTCCCACTCTTTCATGCGCGGCGCAAGGTGGCGCGTAGTTCATCTCCGAGCGCGCT	420	
QY	421	actctgatttgagcagcagctacggtgtgtgatalaatttactttatttcaacggctatg	480	
DB	421	ACTCTGATTGAGCCAGCAGCTACGAGTGTGATATTTTACTTTATTCAACGGCTATAG	480	
QY	481	cgacgaataatcgatgaggttcgagatgctatctacttgcggaggaatcccaatgctggc	540	
DB	481	CGACGAATGATGCGCGGCTTCGATGCTATTCACCTTCCCGAGAGACTCCACAGTGGCG	540	
QY	541	ccaagcgatttaccgctcttcccttcttccacagcatgagagatctccgtaacgat	600	
DB	541	CCAGCGATTATACCGCTTCTCTTCTTCCACACGATGAGGATTCGTTTCACTGAT	600	
QY	601	gaaaggaatctggaacatttagaaggtgaagaaacacctaaggtctcttgatgaacagct	660	
DB	601	GAAGGAAATTTGGAACCTTTAGAAAGTGAAGAAACCTTAAGCTCTGGTGAACACCTT	660	
QY	661	tgatgtcttgagagcctgagtcgctcaagccatgataagtagagatgattgcaatcgg	720	
DB	661	TGATGCTTGGAGCGCTGATGCGCTCAAGGCCATTGATTAAGTACAGATGATGGCAATCGG	720	
QY	721	ggcgtgattcccttcgcatcttcttgacggttaaagatcccttcggaagagcttccgacgg	780	
DB	721	GCGGTGATTCCTTCCCATTTCTTGAGGAGTAAGATCTTGGACAGGCTTTTGGCGG	780	
QY	781	agattcttcgaaagaggtcgaatgacgaacgatttcgctcgaaatgatttgacgaatcc	840	
DB	781	AGATTCTTGGAGAAAGGTGCAATGACGACGATTCGCGAATGATTGAGCAGCAATCC	840	
QY	841	tgcattcttggttgatcagtttcggttcggaagcttcgcttaataagcagaagctcgcaat	900	
DB	841	TGCACTTTCGCTGCTTACGTTTCGTCGGAAGCTTGTAAATACGACGAAGTGCAGAAAT	900	



OY	901	ggaagagatagcaagaggcgtgttagatgttggagagccggttttgttgggtgtgtaagagt	960
Db	901	ggaagagatagcaagaggcgtgttagatgttggagagccggttttgttgggtgtgtaagagt	960
OY	961	aaagaagaaagaaggtatgttaagtgtcagtagagaggtttgaacacagttggggaanaat	1020
Db	961	aaagaagaaagaaggtatgttaagtgtcagtagagaggtttgaacacagttggggaanaat	1020
OY	1021	tgaatcttggtgtctcaatttggaaagtccttcagcaacccttcggtttggaattgttcgtgtac	1080
Db	1021	tgaatcttggtgtctcaatttggaaagtccttcagcaacccttcggtttggaattgttcgtgtac	1080
OY	1081	aaactgagggtggaaatccgcgtctagaaggtatactttccgggtgtccgattgtgtgcctt	1140
Db	1081	aaactgagggtggaaatccgcgtctagaaggtatactttccgggtgtccgattgtgtgcctt	1140
OY	1141	tcgcagatggtgtcgtacaagaagacgaatgtagagcgtcgaatggagagatgtgtgaagacgg	1200
Db	1141	tcgcagatggtgtcgtacaagaagacgaatgtagagcgtcgaatggagagatgtgtgaagacgg	1200
OY	1201	tgttagagttgaagactaaatgaagaggtaagcgttcgtgtgaattgaatgaagagata	1260
Db	1201	tgttagagttgaagactaaatgaagaggtaagcgttcgtgtgaattgaatgaagagata	1260
OY	1261	tattgagaggatgtatggaatgggggagaaagtaagaaacttagagagatgtctgtgcaa	1320
Db	1261	tattgagaggatgtatggaatgggggagaaagtaagaaacttagagagatgtctgtgcaa	1320
OY	1321	gtggaagattgtgcaagaaagcctttggaggaagtgaatctcaagtttaaacactcaa	1380
Db	1321	gtggaagattgtgcaagaaagcctttggaggaagtgaatctcaagtttaaacactcaa	1380
OY	1381	ggtcttcttcgtatgaagtttgtaagtatctaaagacgtlaaattagagtcctccacaaagcaaa	1440
Db	1381	ggtcttcttcgtatgaagtttgtaagtatctaaagacgtlaaattagagtcctccacaaagcaaa	1440
OY	1441	ttgcaaatctcaatctcgttaagttgaatcacttttggccttaatttgttgcagtttgtt	1500
Db	1441	ttgcaaatctcaatctcgttaagttgaatcacttttggccttaatttgttgcagtttgtt	1500
OY	1501	ttcaaatcttctctgttaattttacatttgggtgttaatttagctgtatcttaacgtggaa	1560
Db	1501	ttcaaatcttctctgttaattttacatttgggtgttaatttagctgtatcttaacgtggaa	1560
OY	1561	aataaataatcatctgttggagctcttcacaaatcatcatgtttcccttaattgtctgt	1620
Db	1561	aataaataatcatctgttggagctcttcacaaatcatcatgtttcccttaattgtctgt	1620
OY	1621	aatatataatcatctgttggagactcttcacaaatcatcatgtttcccttaattgtctgt	1680
Db	1621	aatatataatcatctgttggagactcttcacaaatcatcatgtttcccttaattgtctgt	1680
OY	1681	gaaagaagttataatttgcacaaataataaatticaaattcgtccaaatttggaccaa	1740
Db	1681	gaaagaagttataatttgcacaaataataaatticaaattcgtccaaatttggaccaa	1740
OY	1741	gagcgaagatgtctatatttttccacccgaatttccaagagtttggaaattttttttg	1800
Db	1741	gagcgaagatgtctatatttttccacccgaatttccaagagtttggaaattttttttg	1800
OY	1801	caaaatcgactaactcttgacaacaatgcaaaagtataatgataaatttcaaaaaacaa	1860
Db	1801	caaaatcgactaactcttgacaacaatgcaaaagtataatgataaatttcaaaaaacaa	1860
OY	1861	ctcagaacaataaagcccgaaagtatgaataatcaccgtacgnaaatttgcaattgcgc	1920
Db	1861	ctcagaacaataaagcccgaaagtatgaataatcaccgtacgnaaatttgcaattgcgc	1920
OY	1921	tcctattttaattttttttgtgtgttaataaactcgttatcttacttttaataata	1980
Db	1921	tcctattttaattttttttgtgtgttaataaactcgttatcttacttttaataata	1980

QY	1981	aaagcgaatgcataatgcttgcgcgcataatcgaattatgtaattgtaattcagcattt	2040
DB	1981	AAAGGAGATGCATATAGCTTGGAGATATATAGACTTGATGGTAATGTCAGTATTTT	2040
QY	2041	tct 2043	
DB	2041	TCT 2043	
RESULT	2		
AB013596			
LOCUS	AB013596	1507 bp	PLN
DEFINITION	Perilla frutescens PF3r4 mRNA for UDP-glucose:anthocyanin		20-MAR-1999
ACCESSION	AB013596		
KEYWORDS	UDP-glucose:anthocyanin 5-O-glucosyltransferase, complete cds.		
SOURCE	AB013596.1 GI:4115558		
ORGANISM	Perilla frutescens (variety:crispa, cultivar:Shikun) leaf cDNA to mRNA.		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; Gentiananae; Lamiales; Lamiales; Perilla.		
AUTHORS	Yamazaki, M., Gong, Z., Fukushima-Mizutani, M., Fukui, Y., Tanaka, Y., Kusumi, T. and Saito, K.		
TITLE	Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by RNA differential display for plant forms regarding anthocyanin		
JOURNAL	J. Biol. Chem. 274 (11), 7405-7411 (1999)		
MEDLINE	99167509		
REFERENCE	2 (bases 1 to 1507)		
AUTHORS	Yamazaki, M., Saito, K. and Gong, Z.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami Yamazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. & Biotech.; Yayoi-cho 1-33, Inage-Ku, Chiba, Chiba 263-8522, Japan (E-mail:mami@p.chiba-u.ac.jp, Tel:+81-43-290-2905, Fax:+81-43-290-2905)		
FEATURES			
source	Location/Qualifiers		
	1..1507		
	/organism="Perilla frutescens"		
	/cultivar="Shikun"		
	/variety="crispa"		
	/db_xref="taxon:48386"		
	/tissue_type="leaf"		
	17..1399		
	/gene="PF3r4"		
	17..1399		
	/gene="PF3r4"		
	/codon_start=1		
	/product="UDP-glucose:anthocyanin 5-O-glucosyltransferase"		
	/protein_id="BA36421.1"		
	/db_xref="GI:4115559"		
	/translation="MVRRLVLAATFPAGHINPALQFAKRLKAGTDTFTSYAMRBMATVSAAGNPPGDLFVAESDGYDGLKCGGKRYMSMKRGSAALBNLLNNHDMATVYVSHLFAANAAYARASQVSAALMWPATVLCIYYFNGYADEIDAGDEIQLPRLPLEORSLPTFLPETERRLMKKELETLUDGEBKAVLYNTPDALEPALTRIDRELIGIGPLISAFIDGDPSTSGDLEFKSEBNNEVMDLTPKSSVYVYSVSGSVLRFKRAOMEIEIGKGLIACGPFLLMIIRQKNDDEEBEELSTIGELKMKIYVWSCOLEVLAHPALGCEVTHCGMNSAVESISGCVVVAWPQFDOTYAKLLEDAMGAGVVRVNMREGGVGDSIEIRCVEMVMDGSKSLVRENMIKKWTLAREAMGEDGSSLKLNINAPLHOVARA"		
BASE COUNT	366 a	448 g	342 t
ORIGIN	351 c		
Query Match	34.0%;	Score 700.2;	DB 7; Length 1507;
Best Local Similarity	71.3%;	Pred. No. 5e-130;	
Matches 1029;	Conservative	0; Mismatches 365;	Indels 51; Gaps 6;
QY	11	aaaaaaaaaaaaatgacagagctcagctctcttgcgcacatctccacagcagga	70

Db 2 AAAAAATTCACAAAAATGTCGCCGCCGCCGCTGCTAGCAACGTTTCTCCGCCAAGGC 61  
 QY 71 cacataaatccgcgccttcaattcgcgaagcgtctcgaatgacgaatcaatcaatca 130  
 Db 62 CACATTAATTCGCCCTCCATTCGCGAAGAGACTCTTAAAGCCGCGACTGACGTCA 121  
 QY 131 ttcttcacacgcgtctcgaatgacgcgcgtctcgaagacgcgcgtctggtcacaag- 189  
 Db 122 TTTTTCACGAGCGTTTATGATGCGCGCGCATGCCAACACACGCTCCGCGCTGCCGGA 181  
 QY 190 -----gctatcaatttctgtctgtcttccgaagcgtctcgaagcgtctcgaagc 241  
 Db 182 AACCCACCGGGCTTCGACATTCGCGGCTCTCCGACGGCTACGACGCGGCTGAAGCCC 241  
 QY 242 ggaagacatggaagac 301  
 Db 242 TCGGGGCGAGCGGAGCGCTACATGTCGAGATTAAGCCCGCGGCTCCGAGCGCTTAAG- 300  
 QY 302 galactcttgcagcaataatgctcgaatcaaaaagcagcaaaaacacacacacacac 361  
 Db 301 -----AAACCTCTCTTCACACACACGACGCTTCGCTGCTAC 343  
 QY 362 tccacaccttgcagc 421  
 Db 344 TCCACCTCTTTCATGCGCGCGGAGGTGCGCGTGAAGTCCAGGTCCGAGCGCGCTT 403  
 QY 422 ctctgcatgagcgaac 481  
 Db 404 CTCTGGGTGACGCGCGCGCGCGCGCGCTGTGCTATATTACTTCTTACTTCAACGCGCT 463  
 QY 482 gaagaaatcagctcgcgcgtctcgaatcgaatcgcgcgcgcgcgcgcgcgcgcgcgc 541  
 Db 464 GAGGAGATGACGCGCGCGCTTCGAGAAATTCAC---GCTCCCTGGGCTTCACCGCTG 520  
 QY 542 cagcgcgtatccgcgtctcctctcctcctcctcctcctcctcctcctcctcctcct 601  
 Db 521 CACCGGAGTCTTCGACCTTCTGCTGCGCGAGATCCGAGACCGGAGATCCGCTGATG 580  
 QY 602 aaggaagaatctggaatctggaatctggaatctggaatctggaatctggaatctgga 661  
 Db 581 AAGGAGAGCTGGAACATTAGACGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 640  
 QY 662 gatgcgtctggaac 721  
 Db 641 GATGCGGTGAGCGCGCATGACGACGCGTATGATGATGATGATGATGATGATGATG 700  
 QY 722 cgcgtgac 781  
 Db 701 CCGTTGATTCCTTCGCGCTTCTTGGAGCGGAGATCCCTCCGAAACGCTTACGGCG 760  
 QY 782 gatctgctggaagaagcgtcgaatcgaacacacacacacacacacacacacacacac 841  
 Db 761 GATCTTTTCAAAAA---TCGGAGAGAGATTAAGTCCGTGAGTGTGGACACGAGCCG 817  
 QY 842 cgaatcgcgtctggaatctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 901  
 Db 818 AATCTCTCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 877  
 QY 902 gaagaagaatgcaagaagcgtcgaatctggaatctggaatctggaatctggaatctg 961  
 Db 878 GAAGAGATTGGAG 937  
 QY 962 aa-----cgaagaagaagaagcgtcgaatctggaatctggaatctggaatctg 1006  
 Db 938 CAGAGAGATGACAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 997  
 QY 1007 cgaagtggaagaatctgcatctgctcgaatctggaatctcctcgaacacacacacac 1066  
 Db 998 AAATATGGGAGAAATAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1057  
 QY 1067 ggaatctgctggaac 1126  
 Db 1058 GATGTTTCTGACGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1117

QY 1127 cgcgtgtgcttccgc 1186  
 Db 1118 CCGGT 1177  
 QY 1187 gctggaagaac 1246  
 Db 1178 GCGTGGGCGACGCGGTGAGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAG 1234  
 QY 1247 gaattagagatgctatgagagagagagagagagagagagagagagagagagagag 1306  
 Db 1235 GAGATGAGAGAGGT 1294  
 QY 1307 gaagatcgc 1366  
 Db 1295 GAAATATCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1354  
 QY 1367 gtaac 1426  
 Db 1355 CTCAGAAATCTCAACGCGCTTCTTCATCAAGTTGACAGCTGCTTAATACAAATG 1414  
 QY 1427 CCC 1429  
 Db 1415 TCC 1417

RESULT 3  
 AB013597  
 LOCUS  
 DEFINITION  
 5-O-glucosyltransferase homologue, complete cds.  
 ACCESSION  
 AB013597.1 GI:4115560  
 VERSION  
 PF386; UDP-glucose:anthocyanin 5-O-glucosyltransferase homologue.  
 KEYWORDS  
 Perilla frutescens (variety:crispa, cultivar:Shikun) leaf cDNA to mRNA.  
 ORGANISM  
 Perilla frutescens  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 Asteridae; Gentianaceae; Lamiales; Lamiales; Gentianaceae; Perilla.  
 REFERENCE  
 1 (sites)  
 AUTHORS  
 Yamazaki, M., Gong, Z., Fukuchi-Mizutani, M., Fukui, Y., Tanaka, Y.,  
 Kusumi, T. and Saito, K.  
 TITLE  
 Molecular cloning and biochemical characterization of a novel  
 anthocyanin 5-O-glucosyltransferase by mRNA differential display  
 for plant forms regarding anthocyanin  
 JOURNAL  
 J. Biol. Chem. 274 (11), 7405-7411 (1999)  
 MEDLINE  
 99167509  
 REFERENCE  
 2 (bases 1 to 1458)  
 AUTHORS  
 Yamazaki, M., Saito, K. and Gong, Z.  
 TITLE  
 Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Maml  
 Direct Submission  
 JOURNAL  
 Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Maml  
 Yamazaki, Pac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. &  
 Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan  
 (E-mail:mamlyep.chiba-u.ac.jp, Tel:+81-43-290-2905,  
 Fax:+81-43-290-2905)  
 FEATURES  
 source  
 1. 1458  
 /organism="Perilla frutescens"  
 /cultivar="Shikun"  
 /variety="crispa"  
 /db\_xref="taxon:48386"  
 /tissue.type="leaf"  
 29. 1360  
 /gene="PF386"  
 /gene="PF386"  
 29. 1360  
 /codon\_start=1  
 /product="UDP-glucose:anthocyanin 5-O-glucosyltransferase  
 homologue"  
 /protein\_id="BAA36422.1"  
 /db\_xref="GI:4115561"  
 /translation="MVRRRVLLATFPAGQHINPDLQFAKRLKAGTDVTFMFSVAMR



```

/cultivar="Surfinia"
/db_xref="taxon:4102"
/tissue_type="corolla"
100..1506
/gene="PH1"
100..1506
/codon_start=1
/product="anthocyanin 5-O-glucosyltransferase"
/db_xref="GI:6683052"
/db_xref="GI:6683052"
/translation="MGPVHILITTPAOGHINPALQFAKNLVKMGIEVTFSTIAQS
RMDKSIINAPKLNIPFSDGDFDPSKDPVYMSOLRKGSEETVKIILTCSEN
GQITCLLYSIFLPMWAEVAREVHIPSALMSOPATIIDIFYNFKEMANSEN
PMSIOLPGLPLETRDLPFLPYGAGSLVALPPEKLDITDAETPKIIVNTE
DELEPAINATIEGYKFGVIGLIPSAFIGNDPLASFGGLDFTQMSNEMENKPN
SSVYISFGSLMNPISISOMEESKLIIDIGRFLVAKENEGKEEKKLCICELE
KIGKIVPMSOLEVLHPSLGGFVSHGCMNSALESLAGVPAVPQMTDQNTAKQV
EDVMSGVAVRINEDGVVESEIKRCIELVMDGEBKELRNNAKRELAEVAVG
GSSHNKLAFLIDVAKGF"
BASE COUNT      603 a      318 c      369 g      574 t
ORIGIN
Query Match      20.6%; Score 424.2; DB 7; Length 1864;
Best Local Similarity 60.2%; Pred. No. 6.5e-75;
Matches 856; Conservative 0; Mismatches 513; Indels 54; Gaps 7;
14 aataaaaaaaatgagcagagctacgctccttgccacatcccgagcagagacac 73
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
88 AACCAAGAAATATGTGTGACGCTCATGTCTTAACAACATTTCCACACAAAGCCAT 147
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
74 ataaccgccccctcaattcgccaagcgctcgcacaatgcacatccacatc 133
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
148 ATTAATCCAGCACTTAAATTTGCCAAGAATCTTGTCAAGATGGGCAATAGAAGACATTT 207
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 ttcaccagcgctcagcatalgacgcgcatg---tccagaacgcgcgctcgctcaacgag 190
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 TCTACAAAGCATTTATGCCCAAGCCGTATGATGATAAATTCATCTGTATATGCACCAAAA 267
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
191 cctcaaatcttgctgctgttcctgcgcggtatagcaagcggttaagcccgagacgct 280
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
268 GGATTAATTTCAATTCATCTTCCGATGCTTGTGATGAAGGTTTATCATTTCAAAAAGAC 327
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 gggagaagacatactgcgagatgaaagcagagctataaagccttgacgctactcct 310
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
328 CCGTATTTTACATGTCACAACTTCGTAATGTGGAAGTGAACGTGCAAAAAAATATAT 387
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
311 gcagcaataatgctgacatacaaaaagcagcaaaatcacgltcgtagtactccacctc 370
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
388 -----CTCACTTGCTGCTGAAATGGAACGCTATTAACCTTGCTCACTTACTCCATTTTC 441
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 ttgcatgagcgagcagaagtgagcgctgagctccatctccgagcgcgctactcgtgact 430
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
442 CTTCCTTGGGCACAGAGTACGACGTGAAGTTACATCCCTTCTCTCTTTTGAGAT 501
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
431 gaggcagtaagtgctgataatacttacttactcaagggcctatagcagcaaatc 490
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 CAACCGAACATATATGGACATATATTACTCACTTTTCATGATGAAAAAGCTATG 561
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
491 gatcgaggttcgagatgctactctgccc-----ggaagactcccgagctgagcc 541
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
562 GCTAATGAATCCAAATGATCCAAATTTGTCATCACTCCGGGGCTCCACATACGGAA 621
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
542 cagcgatattacgcgtcttccctctcctc-----cgaagcagtagagattc 569
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
622 ACTCGAGATCTCTCTCATCTTTACTTCCATTAGTGTCAAAAGGGAGTCTTCAGTTGCA 681
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
590 cgttcagtgtagaagaatactgaaacttgaaagtgtagaagaataaagcttgcttg 649
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
682 CTTCACCAATTCAAAGATGATAGACATTAATGATGCTGTAACCACTCTTAAGATTCIT 741
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
650 gtgacagccttgatgctgtagcctgtagcgtcacaagggcattgataagtagagatg 709

```

```

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 742 GGAATATACATTTTGAATATAGAGCTGAGGCACTCAATGCAATGAAGCTTATAGTTT 801
Qy 710 attgcaatcggcgctgtgattccctccgcatctcttgagcgtataaagatcccgagag 769
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 802 TATGAAATTTGGACCGGTGATCTCTTCTGCTTCTGGGGTGAATGACCTCTTAGATGCT 861
Qy 770 tcttcgagcagagattgtgctgagaagaaggctgaatgacagatgctcctgaatgctg 829
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 862 TCAATTTGGTGGTGATCTTTTCA-----AATTCAAATGACTATATGGAATGGTTA 912
Qy 830 agcagcaatccctcgatctctcggttgattacgcttcgctcggaagctcgttaacagag 889
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 913 AACTCAAAACCAAAATTCATCAGTGTATTATATCTTTTGGAGTCTATATGAATCATCT 972
Qy 890 aagtcgcaaatgagaagaatgacagaagaaggcgttgatgattgaggagcgcttttgctg 949
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 973 ATTAGCCAAATGAGAGATATCAAAAAGGTTGATAGCATAGGAAGGCCGCTTTTATG 1032
Qy 950 gtgtagaagta-----aacgaagagaagaagtaattgataagcttgcatgag 997
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1033 GTGATAAAGAAATGAAAAAGCAAGAAAGAAAGAAATMAAAGCTTGCTGTATGAA 1092
Qy 998 gagttgaaacgagtgagggaatltgacttgcttccaatlggaagtcctgacgcat 1057
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1093 GAATTGCAAAAAATAGAAAAATAGTTCCATGCTGTTCACAACTTGAACTTCAAAAACAT 1152
Qy 1058 cctcgcttgagatctgctgacacacgtcggttggaatctgcacttaagagttatct 1117
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1153 CCATCTTTAGGATGTTTGTCTCTCATTTGTGATGAAATTCACCTTAAGATTTAGCT 1212
Qy 1118 ttcgggttcctcgatgctgcttcccgagctgctcgaatgcaagagagacgagacgtg 1177
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1213 TGTGGAATGCGCACTTGTGCTCATATGACAGATCAATGAATGACAAATGCAACAA 1272
Qy 1178 atggaagatgctgtagagacgagctgtagagagagacataagagaaggctagctgct 1237
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1273 GTTGAAGATGCTGGAAGAAAGTGAGTAGACATGAATTAATGAATGATGCT---GTTGTT 1329
Qy 1238 gatcgatgaatattgagagatgatttgagagaggttgatgagtgaggagaagaagtag 1297
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1330 GAAAGTAGAGAAATCAAAAAGTGTATTGAATGATGAGAGAGGAAAGGCGAA 1389
Qy 1298 aaacttagagagatgctgcaagtgaaagatttgcaagaagaactatgagagaagat 1357
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1390 GAATGAGAAAGATCTTAAGAAATGGAAGATTTGGCTAGAGAACTGTGAAGGAAGGT 1449
Qy 1358 ggaacttcagttacaacactcaaggtccttcttgatgaggttg 1400
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1450 GGATCTTTCACACAAAGATTTAAAGGCTTTTATTGATGATGCTTG 1492

```

RESULT 5  
 ATCHRIV38/c  
 LOCUS ATCHRIV38 198944 bp DNA PLN 16-MAR-2000  
 DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 38.  
 ACCESSION AL161538  
 VERSION AL161538.2 GI:7268110  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Arabidopsis thaliana  
 thale cress.  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.  
 1 (bases 1 to 198944)  
 REFERENCE  
 1 EU Arabidopsis sequencing project.  
 AUTHORS Direct Submision  
 TITLE Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer  
 JOURNAL Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:  
 Biochemie.mips.biochem.mpg.de/mayer@mips.biochem.mpg.de Project  
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge  
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,

E-mail: michael.bevan@bbsrc.ac.uk  
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.blochem.mpg.de/proj/tbal/> this fragment has an overlap with ATCHNRIV37 at the 5' end and an overlap with ATCHNRIV39 at the 3' end.

/organism="Arabidopsis thaliana"

4845. 5598

```
join(4845. .5031,5330. .5598)
```

```

/note="strong similarity to major latex protein type 1,
Arabidopsis thaliana, contains EST gb:At994471.1, T43693,
At099764, H36765, T43905, At998305.1, R29829, T22625,
R90142, T22170, W43041, R65053, AA712594,
AA712597"

```

```

/codon_start=1
/product="major latex protein like"
/protein_id="CAB78448.1"

```

```
/translation="MATSCTVYVTEPLKSAEKRYKRRNENHLPPDAIGHITQGVTV
HDEWMSHGALIKWNTLDCKPEMKEREIDENMAVFGLEGHWELQKYDTITILE
QFIQKSPDDIYCKITMTWENRADSPSPSYMYKLKSLADMAHVLKA"
1845..5031
```

5032. 5329

5330. .5598

complement(9562. .9654)

complement

	1398..	14072.14674.	14859.15291.	15425))	
/note=	strong similarity to AMP-binding protein - Brassica				
naps;	Contains putative AMP-binding domain signature				
A255-56;	contains ESTR gp:A1999282.1"				

```

/codon_start-1
/product="AMP-binding protein"
/protein_id="CAB78449.1"
/db_xref="GI:7268112"

```

/t/ra1s1a1n-n="MATSS1G1S1T1Y1L1G1S1P1R1O1D1G1F1L1G1S1H1S1V1R1P1R1R1  
 R1E1K1V1K1S1S1P1E1L1S1S1T1G1C1A1L1R1S1E1M1A1D1I1F1A1D1N1S1C1R1M1Y1S1O1G1A1N1V1K1S1R1E1E1  
 L1Y1V1O1J1E1L1D1P1E1G1R1Y1L1G1V1A1D1E1I1A1F1A1D1N1S1C1R1M1Y1S1O1G1A1N1V1K1S1R1E1E1  
 L1Q1Y1R1S1E1S1A1Y1V1D1N1D1E1P1N1I1A1S1P1K1S1L1R1L1L1D1E1K1S1L1Y1O1G1Q1I1V1S1  
 Y1A1I1N1G1E1S1A1L1D1S1N1T1R1S1R1O1F1D1D1T1A1I1N1T1S1E1T1N1G1M1V1H1N1L1N1  
 H1O1H1L1S1Y1P1A1O1A1G1E1F1M1L1S1M1A1E1A1S1Y1F1E1C1G1S1O1M1T1S1Y1L1D1L1K1R1  
 O1P1T1V1S1Y1P1Y1E1L1Y1S1G1I1O1Q1I1S1A1S1A1G1K1L1A1T1L1Y1S1M1A1Y1E1M1K1R1I1B1O1G1L1  
 E1O1K1P1T1A1V1E1V1D1M1L1R1A1T1A1L1M1P1H1M1A1K1L1Y1K1R1H1S1I1G1S1K1I1S1G1S1G1L1  
 H1D1F1E1F1A1G1I1N1G1L1E1T1S1P1V1C1A1L1S1C1N1G1S1G1H1M1E1K1E1Y1D1P1E1N1N1  
 P1P1S1K1I1G1V1O1G1V1A1K1Y1K1P1T1C1O1L1S1E1N1G1F1N1G1D1A1H1S1K1S1R1S1C1H1  
 G1V1L1B1E1R1A1D1T1Y1L1S1T1G1E1V1E1L1E1E1A1M1S1R1V1E1Y1V1G1O1R1R1Y1G1A1I1P1K1E1  
 E1A1Q1V1D1E1T1S1K1E1I1L1Y1E1O1L1K1R1K1T1S1E1C1S1F1O1Y1P1V1L1Y1D1P1T1I1N1G1L1M1P1T1M1K1I1R1  
 01V1A1K1K1E1E1D1O1Y1S1\*

intron	complement(9655. .9749)
exon	/number=1 complement(9750. .9807)
intron	/number=2 complement(9808. .9882)
exon	/number=2 complement(9883. .9998)

```

/number=3
complement(9999. .10096)

```

```
complement(10097. .10176
```

```
complement(10177. .10275
```

complement(10276. .10435

```
complement(10436. .10593
/number=5
```

```
compiletime(10394. :10/31
/number=6
```

```
compnumenc(0) : 00000000  
/number=6
```

/number-7

```

/number-7

```

```

/number=8      11383
count_start+11301

```

```

/number=8
comp]ement/111383 111541

```

```

/number=3
complement(11542. .11652

```

```
complement(11653. .11758
```

complement(11759. .11881

```
complement(11882. .11949
/number-11
```

```
complement(11950.  .12155
/number-11
```

```
complement(1230: 1230-1
/number=12
```

/number=12

```
/number=13
```

```

/number=13

```

```

/number=14
ccmmlomont/13035 13287

```

```

/number=14
comp]ement/13288 13360

```

number-13  
complement(13361). .13957

```
complement(13958. .14072
```

complement(14073. .14673

complement(14674. .14859

complement(14860..15290  
/numbers:17

```
complement(15691. :15423
/number=18
```

```
comp number\1000000 : 1000000  
/number=1
```

/gene="AT4g14080"

```

complete((join(A16137..16174,16252..17604,17687..17732),
/notes- the cDNA A26668 is cognate to this gene; gene A6
for another-specific protein - beta-1,3-glucanase homolog
Contains Glycosyl hydrolases family 17 signature
AA274..287; Prokaryotic membrane lipoprotein lipid
attachment site AA8-18; contains EST gb:AW004339.1,
218400'
/codon_start-1

```











Yr	Accession	Gene	Species	Chromosome	Contig	Fragment No.
1246	Ugaatttggagagatcgtatcgagaggttattgatttgggggagaaagatcggaacttag	1305				
1284	CGAGTTCGGAAGCTGTATGAGATCGTAATGGAGATCGTCAGACGAACGACAGCAATTTGAG	1343				
1306	agaagatgcctgcacagtcggaagattctggcaagaagaagcatcggaagaatgcatcttc	1365				
1344	AAAGAAATGCCCAAAATGGAAGCATTTTGGCTTAAGCAATCAGCAAGCAAAATATGTTCAATC	1403				
1366	agtaacaacctcaaggtctcttcctgactgaagggttga	1402				
1404	TAAATCTTAATCTTAAGGCTTAATGTTAATGATGTTTA	1440				
RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
8	ATCHRIV42/c	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 42.	AL161542	AL161542.2	GI:7268303	Thale cress.
REFERENCE	AUTHORS	TITLE	JOURNAL			
		EU Arabidopsis sequencing, project.				
		Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mps.biochem.mpg.de, mayert@mps.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bpsrc.ac.uk				
COMMENT		Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a> this fragment has an overlap with ATCHRIV41 at the 5' end and an overlap with ATCHRIV43 at the 3' end.				
FEATURES	source					
	exon	1..195068	/organism="Arabidopsis thaliana"	/variety="Columbia"	/db_xref="taxon:3702"	/chromosome="4"
		9917..10066	/gene="AT4g15560"	/number=1		
	CDS	101n[9917..10066,10350..10475,10559..10874,10981..11252,11340..11613,11691..11998,12081..12362,12441..12725,12813..12953]				
		/gene="AT4g15560"				
		/note="GT/AC sites ignored to preserve continuity; light inducible gene required for chloroplast development; contains prokaryotic membrane lipoprotein lipid attachment site AA452-462; prokaryotic membrane lipoprotein lipid attachment site AA591-601; Transketolase signatures AA102-121; Transketolase signatures AA505-521; contains EST gb:AW004332.1, A1997953.1, W43562, N96015, W43388"				
		/codon_start=1				
		/product="DEF (Clai) protein"				
		/protein_id="CAB78598.1"				
		/db_xref="GI:7268304"				
		/translation="MASSAFAFPSTIITNGGLSTDSCKSTSLSSNSLVTLDPSCLIKPNNNSNRRAVCASLAKEGYYSNRAPPLDITINVIHMKLSVKEKLQSLDELRSDVFNKSTGCHGLSSLVEGLTVALIETVPOKLIIMDVGHSGYPIKILTGRRKMTPTMOQTNGLSGFTKRGESEHDCFGGHSSTISAGLMAVGHDLGKNNVAVIGDCAMTGOAAYEAMNNAGYIDMDIVILINDKQVSLPATATDGPSPVGAISALSRQSNPALRELVEAKGRTKQIGCPMHQLAAKVDETARKGISITGSSLFELICITYIGPDGHNIDLVAILKEVKSTRFTGPAVLINHYTEKGRQYPAERADDKYHGVVAFDPATGRQFKTNIDVQSTTTFEALVAEAEDKDVVAIHMAAGGTGINTLQRRFPTRCFVIGIAE				

OHAVTFAGLACDEGLKPCATYSSFMORAYDOVYVHVDLQKLPVRFAMRACGLYAGDG  
PTHCAGFVPTFMACLPNNITVWAPSDSDADLTNNATAYATIDDPSCRTYRNGSIGAL  
PPNGAGVEITIEGKGRILKEGERVALLGYSAVOSCIGAAVMEERGLNVTAVDARCK  
PLDRLILISLAKSHEVLTIVDEGSGIGGSHVQFALDLGLDKLKWPRMVLPPQRYI  
DHGAPADSLAECGLMPSHIATATNLIGABREALF"  
9917. 12953  
gene /gene="AT4g15560"  
10067. 10349  
intron /gene="AT4g15560"  
/number=1  
exon /gene="AT4g15560"  
10350. 10475  
/number=2  
intron /gene="AT4g15560"  
10476. 10558  
/number=2  
exon /gene="AT4g15560"  
10559. 10874  
/number=2  
intron /gene="AT4g15560"  
10875. 10980  
/number=3  
exon /gene="AT4g15560"  
10981. 11252  
/number=3  
intron /gene="AT4g15560"  
11253. 11339  
/number=4  
exon /gene="AT4g15560"  
11340. 11613  
/number=4  
intron /gene="AT4g15560"  
11614. 11690  
/number=5  
exon /gene="AT4g15560"  
11691. 11998  
/number=5  
intron /gene="AT4g15560"  
11999. 12080  
/number=6  
exon /gene="AT4g15560"  
12081. 12362  
/number=6  
intron /gene="AT4g15560"  
12363. 12440  
/number=7  
exon /gene="AT4g15560"  
12441. 12725  
/number=7  
intron /gene="AT4g15560"  
12726. 12812  
/number=8  
exon /gene="AT4g15560"  
12813. 12953  
/number=8  
intron /gene="AT4g15560"  
14413. 14436  
/number=9  
exon /gene="AT4g15570"  
20338. 20373  
/number=1  
CDS /gene="AT4g15570"  
/number=1  
join1.<20338. 20373,20767. 20852,20962. 21012,21256. 21367  
21467. 21572,21653. 21750,21989. 22117,22182. 22274,  
22625. 22678,22957. 23015,23122. 23206,23286. 23765,  
24183. 24269,24366. 24557)  
/gene="AT4g15570"  
note="similarity to SEN1 protein, yeast; Contains  
ATP/GTP-binding site motif A (P-loop) AA13-20; contains  
EST gb:W65040"  
/codon\_start=1  
/product="SEN1 like protein"  
/protein\_id="CAB78599.1"  
/db\_xref="GI:7268305"

exon	20767. . 20852	/gene="AT4g15570"
intron	20853. . 20961	/number=2
exon	20962. . 21012	/gene="AT4g15570"
intron	21013. . 21255	/number=3
exon	21256. . 21367	/gene="AT4g15570"
intron	21368. . 21466	/number=4
exon	21467. . 21572	/gene="AT4g15570"
intron	21573. . 21652	/number=5
exon	21653. . 21750	/gene="AT4g15570"
intron	21751. . 21988	/number=6
exon	21989. . 22117	/gene="AT4g15570"
intron	22118. . 22181	/number=7
exon	22182. . 22274	/gene="AT4g15570"
intron	22275. . 22624	/number=8
exon	22625. . 22678	/gene="AT4g15570"
intron	22679. . 22956	/number=9
exon	22957. . 23015	/gene="AT4g15570"
intron	23016. . 23121	/number=10
exon	23122. . 23206	/gene="AT4g15570"
	23207. . 23266	/number=11

Matches	774	Conservative	0	Mismatches	556	Indels	141	Gaps	10
QY	7	ccaaaaataaataaaaaaatgacagagctcagctccctctgtgcacatccccagaca	66						
Db	4995	CAAAACAATTTCCAACTCTCCACCGGTTCCACACTTTCTATTGTAATTTCCAGCCCA	4936						
QY	67	gggacacataaattccgcgccttcaattcgcgaagcgtctg-----caatgtcgacat	120						
Db	4935	AGGTACATCAACCCATCTCTGAGCTAGCCCAAGCCCTCGCGGAAACATCTGTGTGC	4876						
QY	121	tcaagtcacattcttcaccagcgtctaaagcagtgycgcagatgtccagaacgcgcgtcg	180						
Db	4875	TGAGTCACTTTGCCGCGCTCAATCTCTGCTTACAAACGCCCATGTTCTTACAGAAAA	4816						
QY	181	ctcaaacgsgctcatcaaatlttltgtctgttttcgcagcggtaagcagcgggtltaagcc	240						
Db	4815	CGTCCCGCAACCCATACCTTGCTGACACTACCGATG6CCACAGACAGCGTTTCAATTC	4756						
QY	241	egg-----	276						
Db	4755	CTCTCTTACTCTCGCAAAATCTGTCGAAGAGCCCACTGGAAACTTCATGTCTGATGTAG	4696						
QY	277	aagcagaggtataaaagccttgagagatacctttgcagccaataatgtcgtacaaaag	336						
Db	4695	ACGAGCTGGCAAGAGACACTAACCGAACTAATTCAG-----ATAACCGGAAACAAA	4642						
QY	337	cagcaaatcaacgttcgtgtgtactccacactcttgcattgsgcgcgaagtgycgcg	396						
Db	4641	CAGGCGCTTTACTTGTGCTGTGTTAACACGATTTCTCTCACTTGTGTCGCTGAGACTAGCCG	4582						
QY	397	tgaattcactcgcggaagcgcgtacatctgtgattgagccgtacagtggttggatata	456						
Db	4581	TGAGTTTCACTCTCTTGTGCTCTTCTTTGGGTCCAAACAGTAACAGCTTCTTCATTTT	4522						
QY	457	ttactttatttcaacgctataagcgaagaaatcgatgcgggtctcgatgc-----	507						
Db	4521	TTACATTTATCTCAATGCTGTGACGAAGATGCAATTCACAGATGGTAAATACCCCTTAG	4462						
QY	508	taataactgtcccgagagatcccaagtgctgtgcccagcgtgattacacgtctccctct	567						
Db	4461	TTCTATTAAATTTACCTTCTCTGCACTGTTACTGTCCGTGATATCTCTTTCAATGT	4402						
QY	568	tcctccacgcgtagagattccgtctcaac--tgaatgaaggaataattgaaactttaga	624						
Db	4401	CTCTTCCAAATGCTACGGGTTCTTCTTCAACCCGCTTTCAGAACAGATTTGATTTCACTGAA	4342						
QY	625	aggtagaagaaaaactaagtgctctgtgtgaacagcctttagtgctgtgagcctgaagcgt	684						
Db	4341	GGAAACAAATAAACCTTAAGATCTCTATACACCTTTCAGAGAGCTTGAAGCCAGAACCAT	4282						
QY	685	caagccaattgataagtaagagatgtatgcgaatcgycgcttgattccctccgaattctt	744						
Db	4281	GAGCTCGGTTCCAGATAATTTCAAAATTTG-----CCCTGTC	4245						
QY	745	ggaacgtlaaagaatcccttcggacaaggttcttcgycggagatttgttcgagaagggltcgaa	804						
Db	4244	GGTCGGTTACTTAAGCTTGGAAGCGATTTT-----TCGAG	4210						
QY	805	tgaacgaatgtcctcgaatgtgtgaacgaagaatccctcgatcctcgtgtgtttacglttc	864						
Db	4209	TGCGCGTCAATTCATAAGATGTGTTGGAATCTTAACACCGGATTCGTCGTATTAGTTTTC	4150						
QY	865	gtctggaagcttcgttaatacgaagaagtctgaattgaagaagatagacaagaagsgcttc	924						
Db	4149	GTTCCGGACCGTTGCCGTTGTGAGCAAGAAACAGCTTGTGAGCTTTGTAAGCCGTGAT	4090						
QY	925	agattgtgggaagccgcttlttgtgtgtgtaag-----	959						
Db	4089	ACAAAGTGGAGACCATTTCTGTGTGCTATTACGATTAAGTCGTACAAATAATTAAGAGA	4030						
QY	960	-----taaacgaaggagaagagatctgtataagtttgcataaggaagttgaaacgaagtg	1014						
Db	4029	TGAGCAACAGAGAGAGAGATTTGCATTAAGTATGATTTTCAGAGAAAGCTTCGATGAATAGG	3970						



/note="similarity to UTP-glucose glucosyltransferase; contains\_start=1  
 /codon\_start=1  
 /product="UTP-glucose glucosyltransferase like protein"  
 /protein\_id="CAB78569.1"  
 /db\_xref="GI:7268274"  
 /translation="MKIQVLPSPGICHLSTYELAKRLVRSDDRLMTVIITPPS  
 ISDDVETTYIASLTITASQDRLNTEAISAQAQPDVDEPTVYIEKQKPOVRVAVI  
 KSTGDSERFVAVDMDCSMIDVNEFVPCVMYTSNATLGITLIHQMFDEK  
 PFDVSLSDSEVNELEFPETLRPYPKVCKLPDFTSSDMTLAFARCFRKMYTMTPL  
 VYQKISAAFMVTEGLAVQIETKRFKRGMLVGMGEIYTTVDIEAVYCVMEINSEVRN  
 RKMEAKKHVSAVMDGGSSVALQCKEIDQVTEENV"  
 complement(35222..35273)  
 /number=1  
 complement(35280..35916)  
 /number=2  
 /number=3  
 /gene="AT4g15280"  
 /number=1  
 /number=2  
 /number=3  
 /gene="AT4g15280"  
 /number=1  
 /number=2  
 /number=3  
 /note="strong similarity to UTP-glucose  
 glucosyltransferase; Contains UDP-glycosyltransferases  
 signature AA346-389; contains EST gb:720820, T46737,  
 A1993247.1, A1996826.1, T76043, AA720097, W43072"  
 /codon\_start=1  
 /product="UTP-glucose glucosyltransferase"  
 /protein\_id="CAB78570.1"  
 /db\_xref="GI:7268275"  
 /translation="MKIELVFIPRLGICHLRPTVKLAKQLIGSENRLSTITIIIPSP  
 DAGDSASACIALSTLTSDDRLLHYESISVAKQPPISDPDPAQVYIEKQTKVDAVA  
 ARVDPTKCLASGLTVDVMDCCSMIDVANEFGVCVMYTSNATPLGTMLAHQVMDOKK  
 VYVSELSNVTNELFEPSLRPYPKVCKLPDFTSSDMTLAFARCFRKMYTMTPL  
 ABELPALKMTNELDDLPQVYRPGPVYHLENGDDDEKQSEITLRMTIDQPSKVYFL  
 CEGSLGFEEDQETRETAVALDQSGRLFMCLRHSPNLTQTRDPDRLNLEVEGLEGL  
 ERLDQGLVAAQVAVLEKPAIGETFLVAGWMSITSLMFGVPMTPLPYAQKYNVA  
 ADEMVELCLAVEIKRYLKGDLFAGCEMTYVGAIEDTERAIRRMEDSDVRNNVKEAME  
 KCHPFLMDGSSKALKEKPIQDIENMD"  
 complement(39813..40254)  
 /number=1  
 /gene="AT4g15290"  
 complement(4010139813..40254,40297..40391,40490..40837,  
 40928..41106,41541..41676,41753..41973,42204..42366,  
 42409..42653,42810..43103))  
 /note="similarity to cellulose synthase, Agrobacterium  
 tumefaciens; Contains prokaryotic membrane lipoprotein  
 lipid attachment site AA320-330;Prokaryotic membrane  
 lipoprotein lipid attachment site AA554-564; contains EST  
 gb:A199203.1"  
 /codon\_start=1  
 /product="cellulose synthase like protein"  
 /protein\_id="CAB78571.1"  
 /db\_xref="GI:7268276"  
 /translation="MVPPTADVRESPIITVNTVSLIANYNPANKLIGSYDDGCSB  
 LTFYSLSKSEKVKTIAPPCCKYNNRYVAPRPYFLNPVATDDVSFKDQKMMKITYYV  
 FTVYFCLIMKREYKLCRKVDATGDGSHMDADDDEASNTKPNDSITVYKLKLLK  
 PLKTYVRYVAVQSSKVMYLIKLIIYVWENKGGVDEKVEPHLYVISREKRPYLIHHYTT  
 GANFVLELVNLSHTSEFFYLRMTNAPRYPLANDCCDCAAMEPVRQAMCVLQNSKNRN  
 HCAVFPQPKFYSTYNELAVQSVPLLEFQDFILGRGVAAGQPEYIQTGCFHRR  
 RWYVSGSDSLDNDNGNISQVATREFLADSLVRRKNSKELYKSVYVALQSNKQPS  
 LAILIADAEVGHCHREYQTSKNGNQMNYDSAEEDINTSVSIIHLRGTSSTISDPDPA  
 FISTGYTDLLEAIVQDRMATGAIEVLRNKPQSFVGHGAKTRQFSLAYWALMCALV  
 SIPELlyCLIPAYCLIHDSALPKPCQCLITVYLGMLCLVSLQWPSLGSVSCWML  
 FSIQDIIILKLGISQIEFYIAKRTIPEKRSYKESPSQGEDVRPLNKGKPEPSSGSL  
 FIDGTETLVNINIAALAGYLRVLRSSCSHGGGSGIABACGCIYVMLEFLFKGLPEE  
 HGKYSPLSTLSKAAFLVLYLFFPCVGR"  
 complement(40255..40296)  
 /number=1  
 complement(40297..40391)  
 /number=2

	intron	complement(40392..40489)
	exon	/number=2 complement(40490..40837)
	intron	/number=3 complement(40838..40927)
	exon	/number=3 complement(40928..41106)
	intron	/number=4 complement(41107..41540)
	exon	/number=4 complement(41541..41676)
	intron	/number=5 complement(41677..41752)
Query Match	9.7%;	Score 199.4; DB 8; Length 197419;
Best Local Similarity	52.6%;	Pred. No. 5,3e-30;
Matches 774;	Conservative 0;	Mismatches 556; Indels 141; Gaps 10;
QY	7	ccaaaaaataaaaaaaatgagcagagctcgacltcctcttggccacattcccagaca 66
Db	197414	CACAAACAATTCACACTCTCCACCGGTGCACACTTTCTAATCGCAATATTCCAGCCCA 197355
QY	67	gggaacataaatlccgcgcccttcattctggccaaggtctcg-----caatgcgcaat 120
Db	197354	AGGTCAATCAACCCATCTCTGAGCTATGCGAACAAAGCCTCGCGGGAATAATCTGTGTGC 197295
QY	121	tcaagtcaactcttcaccagcgcgtacgaatgagcgccgatgtccagaacgcgcgtcg 180
Db	197294	TGAGTCACTTGCGCGGCTTCAATCTGTGCCACACCCGCGCATGTTCTTACAGAAAA 197235
QY	181	ctcaaacgggctcatcaatlcttgctgttttcgacgggtatgacgacgggttacagcc 240
Db	197234	CGTCCCCGAAAACCTTAATCTGTGCTAGCTACTCGATGGCCACGACGAGGTTCAATTC 197175
QY	241	cgg-----tagagatgggaagaactcatgctcgaggatgaa 276
Db	197174	CTCTGCTTACTCCGACAAATCTCGTCAAGACGCCACTGGAACTTCATGCTGTGAGATGAG 197115
QY	277	aagcagaggtataaaagccttggagcatactcttgacgcacaaatagtgcgatcaaaaag 336
Db	197114	ACGAGCTGGCAAAGACACTACCCGMACTATTCAGA-----ATTAACGGAAACAAAA 197061
QY	337	cagcaaaaalcacgctcgtgtgttacctccacactcttgcattgvcgvcgcaagttgcgcg 396
Db	197060	CAGGCTTTTACTTGCGTGTGTATACAGATTCCTCTCACTTGGTGGTGGAGCTAGCGCG 197001
QY	397	tgaattccatctccggagcgcgctactctgtgatgtagcagcagctacgtggtgatattc 456
Db	197000	TGAGTTTCATCTCTCTTCTGTGCTCTTCTTGGGTCCACACAGTAAACAGTCTTCTCATTTTT 196941
QY	457	ttactttaattccaacggtatatgagaaatgcatgtcggttgcgtgc----- 507
Db	196940	TTACCAATTACTTCOAATGGCTATGCAAGATGCATATCAGATGCTTAATATCCCCTTAG 196881
QY	508	tattcaacttgcceggagagatccacagttgttgcgccagcggttatctacgtcttccctct 567
Db	196880	TTCTATTAAATTAATACCTCTCTCTGCCACTGTGTACTGTGCTGATTAATCTCTTTTATTTGT 196821
QY	568	tccttcacgcagctlgagagatlcctcgltcac--tgaltaagagagaatctggaaacttga 624
Db	196820	CYTCTTCAATGTCTACGCGCTTCTCTTCTACCCGCTTTCGAGAACAGATTGATTCACGTAA 196761
QY	625	aggtagagaaaaaactaaagttcttggtagaacgctttgatggttggagcctgaaagcgt 684
Db	196760	GGAAGAAATTAACCCCTTAAGATCCTCATACACCTTTCACAGAGCCTTGAGCCAGAAGCCAT 196701
QY	685	caagsgcatitgaatgtagagatgatgcaatcgvgcgcgttgtatcttccctccgaatctt 744
Db	196700	GAGCTCGGTTCAGAAATTAATTTCAAGATTCT-----CCCTGTTC 196664
QY	745	ggaacggtaaagatccttcgacaggtcttctcgcgagagattgttcgagaaaggtcgaa 804
Db	196663	GGTCCGTTTACTTAACGTTTGGAACGGAATTTT-----TTCAG 196629





```
Qy 568 tccctccacgcatgagagatccgttcac---tgatgaaggagaatcggaaacttga 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47347 CTCCTCCAAATGCTACCGCGTTCTTACCGCGTTTCGAGAACAGATTGATTACTGAA 47288
Qy 625 aggtgaagaaataacctaagctctgtgtgaacagctcttgatcgcttgagccctga 684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47287 GGAAGAAATAAACCCTAGATCTCATACACACTTCCAGAGCTTGAGCCAGAACCAT 47228
Qy 685 caaagccattgataagtcgagatgcatgcgaatcgagccgttgatccctccgcatct 744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47227 GAGTCGGTTCCAGATAATTTCAAGATTGT-----CCCTGTC 47191
Qy 745 ggaaggaagaatcccttcgacagctcttcgagcgagattgtctcgagaagggtcgaa 804
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47190 GGTCCGTTACTACAGTTCAGAACGCAATTT-----TCGAG 47156
Qy 805 tgaagacagatgctcctcgaatgtgtgaagcagatccctcgaatcgtgtgttacttcc 864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47155 TCGCGGTGAATACATAGAGTGTGATTAAGCGGATTGCTGCTTATGTTTC 47096
Qy 865 gtctggaagctctgttaataagcgaagctcgcaaatggaagataagaaggctgtct 924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47095 GTTCGGGACGCTTCCGCTGTGACCAACAACACCTGCTGAGCTTGTAAAGCTTGAT 47036
Qy 925 agattgtgagggccgtttctgtgtgtgtaagag----- 959
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47035 ACAAGTCGGAGACACTTCTGTGGTGATTAACGATTAAGTCGTACAGAAATAAAGAA 46976
Qy 960 -----taaacgaaggagaagagatgataagtcgacggaaggagtgtaaacagtgagg 1014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46975 TGAGCAAGAGAGAGAGAGAGATTCATACATAGATTGTCAGAGAACAGCTCGATGAGATAGC 46916
Qy 1015 gaataatgatactgtgtgtcttcaatgtgaagctctcgaatccctcgtgtgagatgtt 1074
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46915 AATGCTGCTTACATGATGATGATCAAGTTAGGTTTGAATCATAGATGATAGATGATGTTT 46856
Qy 1075 cgtgacacacacgctgtgtgaatctcgaactcgaagatgatactctcgtgtgtcgaatgt 1134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46855 CGTGACGCAATTGCGGGTGGAACCTGACGCTGAGAGCTGTGTTTCAAGAGATTCGCGTGT 46796
Qy 1135 ggccttctcgaatgtgtcgaatcaaggagcgaatgcaagctgtatggaagatgtgtgag 1194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46795 GCGCTTTCCGAATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 46736
Qy 1195 gacgggtgtgtgagatgtgaagcgaatgaagcgaatgaagcgaatgaagcgaatgaag 1245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46735 AACAGGTGTAGAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 46676
Qy 1246 tgaactgaagagatgataatgagagatgagatgagagagagagagagagagagagagag 1305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46675 GGAGATACGGGGGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 46622
Qy 1306 agagagatgtctgtgcaagtggaagatctggcaagaagaagcgaatggaagagatgtgacttc 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46621 AGGAATATCCACGAGTGGAAGATTTAGCGCGAGCGCTGTGAGAGAGAGAGAGAGAGAGAG 46562
Qy 1366 agttaaacacactcaagctcttcttgatgag 1396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46561 CTTTATCATCTCAAAAGCTTTGTGATGAG 46531
```

```
RESULT 11
LOCUS ATU01293 1589 bp mRNA PLN 02-JUN-1997
DEFINITION Arabidopsis thaliana UDP-glucose:indole-3-acetate
beta-D-glucosyltransferase (laglu) mRNA, complete cds.
ACCESSION U01293
VERSION U01293.1 GI:2149126
KEYWORDS thale cress,
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
euhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE
AUTHORS 1 (bases 1 to 1589)
TITLE DNA Sequence of UDP Glucose: Indole-3-acetate
        Beta-D-glucosyltransferase from Arabidopsis thaliana (accession No.
        U01293) (PGR97-044)
JOURNAL Plant Physiol. 113: 1004 (1997)
REFERENCE
AUTHORS 2 (bases 1 to 1589)
TITLE Thornburg, R.W. and Graham, R.A.
        Direct Submision
JOURNAL Submitted (06-DEC-1996) Biochemistry and Biophysics, Iowa State
        University, Ames, IA 50011, USA
FEATURES
source
    1..1589
        /organism="Arabidopsis thaliana"
        /variety="Columbia"
        /db_xref="taxon:3702"
        26..1589
        /gene="laglu"
        26..1450
        /gene="laglu"
        /EC.number="2.4.1.121"
        /function="transfers the glucose moiety from UDP-glucose
        to indole-3-acetic acid to form ester linked IAA-Glc"
        /note="IAA-glucose synthase"
        /codon_start=1
        /product="UDP-glucose: indole-3-acetate
        beta-D-glucosyltransferase"
        /protein_id="AA058497.1"
        /db_xref="GI:2149127"
        /translation="MANNNSNPSPGPHFLPVTPEAGHINPISLEAKRATGISGARV
        TPAASIVAVNRMRSPENVEPTLIFASYPGHDDGFEKSAVSDKSRDADGNMSEMR
        RKQKETITETIEDNRKONRPFCTVYITLLTVAELAREHLRSLALVQPVYFSEI
        YHYFNGEDIAISEMANTPSSIKLPSLPILTVRDPISFIVSNVYATLPAFRQIDS
        LKEINPKILINIFOLEPEAMSVDPNFRIVPGPLTVRDFSSRGEYIEMDTKA
        DSSVLYSFEGLTALVSKQLVEICKALIGSRPFLVITPKSYRNKDEQEKEDCIS
        SSKSFPEIGMIVVSWCDQFVNLNHRISIGCVTHCGMNSTLESIVSGVVAFPQMNDO
        MTNALKLEDDWKTGVRVMEKKEEGVYVVDSEIRICRIEEVMEKAEFFGNATRKMD
        LAEAVREGSSFRHLLKAFVDEH"
        1568..1575
        /gene="laglu"
        1588..1589
        /gene="laglu"
        /note="24 A nucleotide"
BASE COUNT 433 a 342 c 375 g 439 t
ORIGIN
```

```
Query Match 9.6% Score 197.8; DB 8; Length 1589;
Best Local Similarity 52.5% Pred. No. 1.1e-29;
Matches 773; Conservative 0; Mismatches 357; Indels 141; Gaps 10;
Qy 7 ccaaaaaataaaaaaataatgagagagcctcgtctcttggcaccatccagacaca 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 CAACACAAATTCACAACTCCACACCGGTCACAACTTCTATTCGAACATTTCCAGCCCA 90
Qy 67 gggacacataaaccgcctcctcaatcgcaagcgctcg-----caatgcgcacat 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 AGGTCACATCAACCCATCTCGAGCTTAGCCAAACGCTCCCGGACAACTCTCGGTGC 150
Qy 121 tcaagtcacatcttcacacagcgtctacgacgtgscgcacatgctccagaacgcgcgcg 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 TCGAGTCACTTCGCGCGCTCAATCTCTGCTTACAAACCGCGCATGTTCTTACGAAAA 210
Qy 181 ctcaaacagggtcacaatcttgtgtgtcttcgcagcgggtatgacgagcgttaccagcc 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 CGTCCCGAAMACCTTAATCTTCGTACTACTCGATGGCCACAGCAGCGGTTTCAATC 270
Qy 241 cgg-----agacgagtggagaactacatcgtcggagatgaa 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 CTCGTCTTACTCCGACAAATCTCGTCAAGACGCGCACTGGAACCTTCATGATGATGAG 330
```



QY 277 aagcagaggtataaagacgtttagcgatactcttgcagccaataatgctgataaanaag 336  
 DB 331 ACACGCTGGAGAAAGACACTAACCAGAACTTAATCGAAG-----ATAACCGGAACAAA 394  
 QY 337 cagcaaatcaacgttctgtgtgctgactccaccccttctgcataggcgccaaagtgagcg 396  
 DB 385 CAGCGCTTTTACTTGGGTGGTTTACACGANTTCCTCACCATTGGGTGCTGACGTAAGCGCG 444  
 QY 397 tgaattccatctccggaagcgctactcttgatlttagccagctacggtgtgttatatatt 456  
 DB 445 TGAGTTTCAATCTTCTGCTCTCTCTTGGGTCCAAACGTAACAGTCTTCTCCATTTT 504  
 QY 457 ttaactttatctaacgagcatagcgaaatcgatgagcggtgctgagtc----- 507  
 DB 505 TTACCAATCTTCAATGAGTACGAAGATGCAATTCAGAGATGGCTTAATACCCCTCTAC 564  
 QY 508 taticacttgcggaggaacccagctgctgagccagcgctgattaccgcttcttct 567  
 DB 565 TTCTATTAAATTAACCTTCTCTGCTGCTTACTGTGCTGATATTCCTTCTTCAATGT 624  
 QY 568 tctctccagcatgagagatccgttcaac--tgatgaagagaatltgaaacttaga 624  
 DB 625 CTCTTCCAAATGCTACGCGCTTCTTCTACCCCGGTTTCGAGAACAGATTGATTCACAGAA 684  
 QY 625 aggtgagaagaaaacccaaggtctgtgtgacagctttagtcgcttgagagctgagtcgct 684  
 DB 685 GGAAGAAATTAACCTTAAGTCTCTCAATCAACACTTTCACAGACTTCAAGCAAGCCAT 744  
 QY 685 caagcgcatatgaatgaacagatgtatgcaatccgagcgcttgatctccgcatctt 744  
 DB 745 GACCTGGTTCCAGATTAATTTCAAGATTGT-----CCCTGCTC 781  
 QY 745 ggaaggttaagatccttcggacaggtccttcggcgagagatltgtcagagaagagtgaa 804  
 DB 782 GGTCCCTTACTAAGCTTGAACGGAATTTT-----TCGAG 816  
 QY 805 tgcagcgatctgctcgaatggttgcagcaatccctcgatcttcggtgtgttgcgttct 864  
 DB 817 TCGCGGTGAATTAATGAGGTGGTGAATTAACGCGGATTCGCTGTGTATTAATTTTC 876  
 QY 865 gtccggaagcttcgttaatacagcgaagtcgcgaatgaagagatagcaagagcgctgt 924  
 DB 877 GTTCGGGAGCGCTTGCGGTGTGAGCAAGAACGCTTGTGAGACTTTGTAAACGCTTGAT 936  
 QY 925 agatgtgagagcgcttctgtgtgtgtgaag----- 939  
 DB 937 ACAAGTTCGAGACCAATTTCTGTGGGTGATTACGATTAAGTCGTACAGAAATTAAGAGA 996  
 QY 960 -----taaacgaagagaagaagatttgaatgtcagatgagagagtgaaacgagtgg 1014  
 DB 997 TGAGCAAGAGAAAGAAAGAAATTCATTAAGTACGATGAGAAAGCTTCATGAGATGAG 1056  
 QY 1015 gaaatgtatctgtgtgttctcaattggaagtcctgaagcattccctcgttggagattt 1074  
 DB 1057 AATGTGTGTTCTAGTGTGATCAATTTAGGTTTGAATCATAGATCGATAGCTGTGTT 1116  
 QY 1075 cgtgaacacgttgcgggtggaatctcgaactcagagagatatacttccggttctcgatgt 1134  
 DB 1117 CGTAGCAGCATTTGCGGGTGAACCTCTACGCTGAGAGAGCTTGTTACAGAGTTCGGGTGT 1176  
 QY 1135 ggccttcccgagtggttgcatacaaggaagcaatgcgaactcgtatgagagatgtgtgag 1194  
 DB 1177 GGGGTTTCCCAATGGAATGATCAGATGAGAAACCGCAACCTTTTACAAATTTGTTGAA 1236  
 QY 1195 gacgggtgtgagtgagagctaatgagga-----gggtagcgctgtttagtga 1245  
 DB 1237 AATAGGTGAAGTGTGAGAGAGAAAGAAAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGT 1296  
 QY 1246 tgaatgaagagatgtatgtgaagaggtatgtgagtggtgtgtgtgtgtgtgtgtgtgtgt 1305  
 DB 1297 GGAATATACGCGCGGTGATTTGAGCAAGATTATGGA-----AGACAAGCGGAGAGGATTAG 1350

QY 1306 agagagtgctggaagtggaagatttggcaagaagaaactatgaggaagatgattctt 1365  
 DB 1351 AGGAATGCGACAGAGGAGGAGGATTTAGCGCGAGAGCTGTGAGAGACGAGGCTTTC 1410  
 QY 1366 agttaaacaacatcaaggtcttctttagag 1396  
 DB 1411 CTTTAATCATCTCAAGACTTTTGTGATGAG 1441

## RESULT 12

AC007153/103223 bp DNA PLN 17-MAY-1999

Arabidopsis thaliana chromosome I BAC F3120 genomic sequence,

complete sequence.

AC007153 AC007153.2 GI:4580365

## KEYWORDS

## SOURCE

## ORGANISM

thale cress.  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.

1 (bases 1 to 103223)

## REFERENCE

## AUTHORS

Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
 Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,  
 Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,  
 Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.S.,  
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.

## JOURNAL

## REFERENCE

## AUTHORS

2 (bases 1 to 103223)  
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
 Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,  
 Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,  
 Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.S.,  
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.

## JOURNAL

## REFERENCE

## AUTHORS

3 (bases 1 to 103223)  
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
 Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,  
 Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,  
 Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.S.,  
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.

## JOURNAL

## REFERENCE

## AUTHORS

4 (bases 1 to 103223)  
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
 Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,  
 Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,  
 Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.S.,  
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (08-APR-1999) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (17-MAY-1999) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA

## COMMENT

On Apr 9, 1999 this sequence version replaced gi:458084.  
 Bases 1-33654 of clone of F3120 overlap with bases  
 50550-84203 of 'TAMU' clone T25N20, gb|AC005106.  
 e-mail for correspondence: arabidsequence.stanford.edu  
 Genes with similarity to proteins in the databases are described as  
 'putative', '-like' or 'similar to'. Genes that have EST  
 similarity but no significant protein similarity are described as  
 'unknown proteins'. Genes that are annotated based only on gene  
 prediction software are described as 'hypothetical proteins'.  
 The software programs used to predict genes include: Grail  
 (informatics group, Oak Ridge National Laboratory,  
 http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge,  
 http://genomic.stanford.edu/~chris/GENSCANW.html), Fexa (V. Solov'yev



QY	857	tacgcttcgcttcggagaagcttcgttaatacagcaagagtcgcaatcggaaagatagcaaga	916
Db	83279	TATGATATCATTTTGGAGCGCTGGTGGCTCTCAAAAAGATCAACTGATTAACATAGCGGGC	83220
QY	917	ggcgcttcagatcttcggagagcgctttcttctgtgtgaagataagatacgaagagagaag	976
Db	83219	GGCTGCAACACAGCGGACATTTCTTTTGTGGGTTGTAGAGAGACGGAGAGAGAAAA	83160
QY	977	glatcgtataagcttcgcatggagagcttcgaacagagtcgggagaaatctcatcttgctct	1036
Db	83159	CTTCCAGAAAACATATATAGAGAAATTGGGTGAGAAAGACGTACCGGTAGCGTGAATCCA	83100
QY	1037	caattcggaaagcttcgcagcaatccctcgttcggagtgcttcgtgcacacatcgcggtgaat	1096
Db	83099	CAACTTGAAGTCTCTACACATATAATTCGATCGGTTGGTTGGTGACACATTGTGGATGAAC	83040
QY	1097	tcgactctagagaagcatatccttcggggttcgcgatagcttggtcttccgaaagtgcgt	1156
Db	83039	TCCAGCTTAGAGAGGATGTGATGTTGGAGTCCAAATCATTTGGTATCGCTATTGGCGCAGAT	82980
QY	1157	caagggagcaatcgccaaagcttgaatggagatgctgtcgtgagagacgggtctgaagctgaagact	1216
Db	82979	CAGCCTACAAATGCTAATAGTTCAATGAGAGATGCTGTGAAGTTGCGATTTAGGTTAAGC-	82919
QY	1217	aatggagaggttagcgcttcgttcgatcgtgataatgaatgaagatgtaattgaagagctatc	1276
Db	82920	--AGACAGATGATGGGTCTGTGGAAGACAGAAAGTTTGTGAGACGTTGTGGAAGAGCTTATG	82863
QY	1277	gctgggggagaaagatgaagtaaaccttaagagagtgctcgcaagtcggaagattggca	1336
Db	82862	GA---GGCAGACCAAGGTAAGAGATTAAAGAAAGATCGAGCAAAATGGAAAGTGGGCT	82806
QY	1337	agaaaaagctatggagaaagatgtaattcagcttaacaactcaagctctcttcgtatgag	1396
Db	82805	CAAGAGGCGTCTTCGAAGAGAGTGTCTGATPAAGACATCAATGAGTTTGTCTTATG	82746
QY	1397	gctc 1399	
Db	82745	TTT 82743	

RESULT	13			
LOCUS	AC006553			
DEFINITION	AC006553	99188 bp	DNA	PLN
ACCESSION	AC006553			05-APR-2000
VERSION	AE002093			
KEYWORDS	AC006553.7	GI:6598600		
ORGANISM	HTG.			
SOURCE	thale cress.			
REFERENCE	Arabidopsis thaliana			
AUTHORS	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
	1 (bases 1 to 99188)			
	Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Feldblyum, T.V., Mason, T.M., Bowman, C.L., Barnstead, M.E., Fujii, C.Y., Mason, T.M., Bowman, C.L., Lee, J.J., Ronning, C.M., Koo, H., Moffatt, K.S., Cronin, L.A., Shen, M., Vanden, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Cressy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.			
TITLE	Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana			
JOURNAL	Nature	402 (6763),	761-768	(1999)
MEDLINE	20083487			
REFERENCE	2 (bases 1 to 99188)			
AUTHORS	Lin, X.			
TITLE	Direct Submission			
JOURNAL	Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA			

COMMENT On Dec 17, 1999 this sequence version replaced gi:4887740.  
The sequence and annotation of chromosome 2 were merged from those  
of the individual clones on this chromosome after removing  
overlaps. For detailed information, please see the TIGR web site  
(<http://www.tigr.org/tcdb/at/at.html>).

Genes were identified by a combination of three methods: Gene  
prediction programs including GRLIL  
(<http://arthur.epm.ornl.gov/pub/xgrail>), Genefinder (Phil Green,  
University of Washington), Genscan (Chris Burge,  
<http://genome.stanford.edu/GENSANM.html>), and NetPlantGene  
(<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the  
complete sequence against a peptide database and plant EST  
databases at TIGR, and manual curations based on those analyses.  
Annotated genes are named to indicate the level of evidence for  
their annotation. Genes with similarity to other proteins are named  
after the database hits. Genes without significant peptide  
similarity but with EST similarity are named as 'unknown' proteins.  
Genes without protein or EST similarity, that are predicted by two  
or more gene prediction programs over most of their length are  
annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
predicted by tRNAscan-SE (Sean Eddy,  
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were  
identified by RepeatMasker (Arian Smit,  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are  
numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones  
F6P23, F5f6, T17A5, and T13J16, the ESSA group for sequencing clone  
F13D4, and Scott Jackson, Jimmy Jiang, Klaus Meyer, Eric Richards  
and Satoshi Tabata for helpful assistance. In addition, we would  
like to thank the TIGR Bioinformatics Department, especiallyixin  
Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Jiang, Jeremy  
Peterson, Michael Holmes, and Delwood Richardson for software and  
database support.

This work was supported by the National Science Foundation,  
Department of Energy and the US Department of Agriculture.

FEATURES	Address all correspondence to: <a href="mailto:at@tigr.org">at@tigr.org</a> .
source	location/Qualifiers 1..99188 /organism="Arabidopsis thaliana" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="11"
misc_feature	<1..1802 /note="Sequence from clone T9H9" complement(<1064..>2005) /gene="At2g31720" complement(<1064..>2005) /gene="At2g31720" /note="T9H9.24; predicted by genscan" complement(1064..2005) /gene="At2g31720" /note="hypothetical protein" /codon_start=1 /protein_id="AAD32277.2" /db_xref="GI:6598601"
mRNA	/translation="MVRGDDQLNGLGGKRALDILCVATERTVSFAVEEIGETSTEGS DKRRERLRRLINYSLTSMNSLSRKLIKIDCKNDPEONNRVAASSSSCHLESS KPQKVYVKNKRPRRPVIIVDEDRTPTEMLIDVARVNGMDAKLIFVKYLPNSDVSDG LOFLAMPWKOIILDMDFLNFEELKEIDRHNNKSASQSGADVIYVNSKGQRKLKLRK WMDSITNYVLGICMNKYVDNIIOGRTSLRIWSFSPDLFFAFVLSDDPAPEELL NLAKLCBETTCEALDPANRMSSLVSDTELDELRL"
CDS	complement(1803..>99188) /note="Sequence from clone F20M17" complement(join(3588..3973,4058..4094,4189..4265)) /gene="At2g31730" complement(3588..4265) /gene="At2g31730" /note="F20M17.23" complement(join(3813..3973,4058..4094,4189..4260))
gene	
misc_feature	
mRNA	
gene	
CDS	

[illegible]

LOCUS	AC005106	84203 bp	DNA	PLN	16-DEC-1999
DEFINITION	Genomic sequence for Arabidopsis thaliana BAC T25N20 from Chromosome 1, complete sequence.				
ACCESSION	AC005106				
VERSION	AC005106.2	GI:6587719			
KEYWORDS	HTG.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphylloids; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.					
REFERENCE	1 (bases 1 to 84203) Dewar, K., Feng, J., Kim, C., Li, Y., Sun, H., Shinn, P., Buehler, E., Kurtz, D., Oji, O., Shen, Y. K., Toriumi, M., Conway, A., Conway, A., Kurtz, D., Oji, O., Shen, Y. K., Toriumi, M., Vysotskai, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and Ecker, J. R.				
TITLE	Genomic sequence for Arabidopsis thaliana BAC T25N20 from Chromosome 1				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 84203) Ecker, J. R.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	3 (bases 1 to 84203) Ecker, J. R.				
TITLE	Direct Submission				
JOURNAL	Submitted (31-JUL-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	4 (bases 1 to 84203) Ecker, J. R.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	5 (bases 1 to 84203) Shinn, P., Brooks, S., Buehler, E., Chao, O., Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bel, B., Chin, C., Chou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Muhsarsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A. and Ecker, J.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-DEC-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA				
COMMENT	On Dec 16, 1999 this sequence version replaced gi:3366536.				
FEATURES	Location/Qualifiers				
Source	1..84203				
	/organism="Arabidopsis thaliana"				
	/db_xref="taxon:3702"				
	/chromosome="1"				
	/complement(join(1491..2159,2389..2452,2568..2703,3061..3149,3231..3277,3369..3457,4022..4139,4246..4353))				
	/note="5' end of hypothetical protein Y0P8H12.1; overlaps T7 end of Y0P8H12"				
	/codon_start=1				
	/evidence-not_experimental				
	/product="T25N20.1"				
	/protein_id="AAC80580.1"				
	/db_xref="GI:3935137"				
	/translation="MGKEKQKDHSHSVESDKEAVHLKKHSEPLTKOEKFCNRACVGRFLRIKGDVNVKAAQRLRSCLSWRSSLGIESLADFEELAGLAVYGLDDETRPVLVRIKQDYKQKLTOKQLTRLVLTLEVAISTSRNVEQFVILDAIAVALVLTILVSEFSNPLILICNSTANLPPEFFFIYKTNKLOISISFHEHCKQTONSRGILRTFEDLSTAMIVSMODPDSFDYDSSSPRVSLSRDTSSLSKTDGKISGASSRAPIVSDGDLTYKPKCLTLTDTSTSLGHTGAYISPLNARSFSPSPARSPFGGPRNSFSTSPMPARTDTRHSIGLADRPISPPSPALFFRSHSVKSEKPRDSFVQELKFRARYDEVTYRSKMRPPIGLGGLVSVSTQIRRRHVSLSGRF"				
CDS	join(9304..11268,11387..11513,11607..11692,12103..12341,12698..12784,13257..13356,13451..14263)				
	/note="unknown: similar to Ests dbj c26511 and gb A1054650"				
	/codon_start=1				
	/evidence-not_experimental				
	/product="T25N20.2"				
	/protein_id="AAC80581.1"				
	/db_xref="GI:3935138"				
	/translation="MEGGGSGGERSRLASNTPRSLKEKSDGRTVSTSTERNYOKRSMTVSDSESDERFMKPPRRSGYDKRTLAKKEKTVKRBDYVHDNRKYVRRNNDASGSPMNRKLTIDFEDEIDGDSANLRRKRFNDGSGVGRGSRFASRDVSGRSGSGR EGLFDRRNRITVNGTICASQSDSSSESDSPMRVQDINGVLKYVNNKNTLAASIKPRDAEITERPSSSRKQRRNNVYKPPFRSGSNVNDNNESESDMSKSRKSEYSKKPEEFTKSTFPELVNPVYRREERKRGSDGDKQLRIRIKGMLDAMTIDYKRRVSDYAAVYVNSGTAVYSITRAYDMLKOLDEGDAPRRDITAAVASEETVYKLARKKTRESEMTKKNKQSSGDSSEKSPGSAVYTPDSEERISLIGKSKRKNAGADDELHAKSKSLTYNNRSPCGSDSHLHGRKTKITRCTLNRSKDKKNPALNGFNPSGKRTLMSLIESGVVQLKQVYKRRKAKYMLEGMITREIHCDCSKILTYSREIIRHAGSKSCQPFONILIESGASLICOVARNMKNQDATNIALHQVPTDNDPDDAGCIGCGDGLICDPCDPSYTHQNCIGMVGVLPSGDHPCNCTKPCDAVAASGKDGNFISLSCGCERRYHQLCLNDEAHKVVSRGASSPGCEKCLEFEKLOLYGVKTEIEGYSWLIHRYVDPDPSFNSQMSAQRITENSKSLAVGLAIDCEPLIVDRSSCYDLIRNYLXGCSNFRNRYTGFYTAILERDETISAASLRHFGKQLAMPETIGTRHITRROGMCRLFPAISAMSKLVERKVLPAIPALDITAHVNGEFLPDDSVRKEKSLNLTLYPPGIDMLKPLHBEENILVAVETAMGTICKPDELSRHRFGSEGISSPQITLKSGLKAVGHGICDDTSSCEGDELTVNVEADSLKLSOELIOOASAFKVENNLSLSISGRGSSDLSISQGVKSEORTSSNLDYVPSCKQNIILVPGAKLMDKSDARADPIL"				
	complement(join(14820..14957,15038..15420,15643..15721))				
	/note="hypothetical protein"				
	/codon_start=1				
	/evidence-not_experimental				
	/product="T25N20.3"				
	/protein_id="AAC80582.1"				
	/db_xref="GI:3935139"				
	/translation="MGFLVAVMNPSPYLVAHHMKSRCPOCKNEKLROGOTSLLFDRGRFLKCAVGAASFPAITIEFSGIQAASEKILDEDEEGVAGVETLFDENERKSEKLPKAYLKSAREVYKTRRESLKENPKDNKRRFADAKSIRYLLNMWGOKTVAGEESVELENVIRALAKFYSKGPSAPLDEYKTEILDLIDNAEEL"				
	join(16409..16661,16979..17133,17234..17258,17346..17362)				
	/note="unknown: similar to Est emb 226799"				
	/codon_start=1				
	/evidence-not_experimental				
	/product="T25N20.4"				
	/protein_id="AAC80583.1"				
	/db_xref="GI:3935140"				
	/translation="MMMDPPRTKEPALAVYVYDMKRCPPVDYDARVRPCTIKRILLKSGNGVNTITAVGSLSKVPRLILEVSSGISLGHVAPLEDEESECSETGPASWICSVCRICGPIAGGVQDNFITHLSRHEHLEKORFTNGHLAN"				
	complement(join(17577..18026,18231..18796,18981..19410))				
	/note="similar to RNA polymerase II elongation factor homolog (S59773)"				
	/codon_start=1				
	/evidence-not_experimental				
	/product="T25N20.5"				
	/protein_id="AAC80584.1"				
	/db_xref="GI:3935141"				
	/translation="MFLHYVLEAASISFYIECSYNNDSINDIQSNIGSEPHILLRPVSPLESSEGLYAPENMNPEDDTWMKAGPISIGLGHVADRYLTPPKLPLDIEILRNKVFERSLSQRYIRVHPADVQKFPASFSWSIPCDHGGGVLPQKQVLPVYSSDEPMRDGSDTAVKAGNEKSCRSLMPCDIEFLPAPCDICGGERKFCVDCCLICCKLISLEHGYSYIKCEAVVSEGHICGHVAMNCLALRYLAGTIGGSGGLDEYECRCKAKIDLEPHYKFLLEICQTVYOGGDVETKLTILRGAGMRNAELNIGESTYIKITFILVYVSDSGEARENDTLOSLODYTPICGPIPNFHEAMKRLHEEICGYDLAKRAQEFYQIAGEKLLAQKECLSDYLQLEKEKSELSRVSQTDANSLMTNVLKRLDQIRKEYTKLMEHVEVAGFGFTPGVIEEYFHLIED"				
	22098..22826				
	/note="unknown: similar to Ests emb 233701 and emb 234709"				
	/codon_start=1				
	/evidence-not_experimental				
	/				





**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:17:13 ; Search time 236.66 Seconds

(without alignments)  
2179.904 Million cell updates/sec

Title: US-09-147-955-5

Perfect score: 2062  
Sequence: 1 atttcccaaaaaataaaa.....taaaaaaaaaaaaaaaaaa 2062

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2062	100.0	2062	1	X02828
2	698.6	33.9	1506	1	W09905287 Seq ID 3
3	680	33.0	1474	1	W09905287 Seq ID 1
4	484.4	23.5	1671	1	W09905287 Seq ID 2
5	424.2	20.6	2105	1	W09905287 Seq ID 6
6	178.4	8.7	1513	1	W09905287 Seq ID 4
7	147	7.1	2062	1	W09905287 Seq ID 3
8	143	6.9	566	1	W09905287 Seq ID 3
9	141	6.8	1731	1	W09905287 Seq ID 3
10	135	6.5	568	1	W09905287 Seq ID 3
11	96.4	4.7	1624	1	W09905287 Seq ID 3
12	96.4	4.7	1626	1	W09905287 Seq ID 3
13	96	4.7	1578	1	W09905287 Seq ID 3
14	83.4	4.0	432	1	W09905287 Seq ID 3
15	66	3.2	1597	1	W09905287 Seq ID 3
16	59.6	2.9	1607	1	W09905287 Seq ID 3
17	52.2	2.5	1671	1	W09905287 Seq ID 3
18	50	2.4	58407	1	W09905287 Seq ID 3
19	49.8	2.4	4810	1	W09905287 Seq ID 3
20	49.6	2.4	794	1	W09905287 Seq ID 3
21	49	2.4	9789	1	W09905287 Seq ID 3
22	48.8	2.4	2750	1	W09905287 Seq ID 3
23	48.4	2.3	1511	1	W09905287 Seq ID 3
24	48.4	2.3	1511	1	W09905287 Seq ID 3
25	48.4	2.3	1511	1	W09905287 Seq ID 3
26	48.4	2.3	1511	1	W09905287 Seq ID 3
27	48.2	2.3	4838	1	W09905287 Seq ID 3
28	48	2.3	405	1	W09905287 Seq ID 3
29	47.4	2.3	940	1	W09905287 Seq ID 3
30	47	2.3	4590	1	W09905287 Seq ID 3
31	47	2.3	6766	1	W09905287 Seq ID 3
32	47	2.3	6768	1	W09905287 Seq ID 3
33	47	2.3	6768	1	W09905287 Seq ID 3

34	47	2.3	8457	1	066797	Ampev spheroidin g
35	46.6	2.3	110000	1	X20248_06	Continuation (7 of
36	46.6	2.3	111309	1	X20250	Borrelia burgdorfe
37	46.2	2.2	110000	1	X20248_00	Borrelia burgdorfe
38	45.8	2.2	5852	1	Q11710	Dictyostelium plas
39	45.6	2.2	4821	1	T38305	Tosporivus M-RNA s
40	45.6	2.2	5760	1	N50530	Sequence encoding
41	45.4	2.2	19124	1	T72882	Plasmodium var-7 g
42	45.4	2.2	110000	1	X20248_05	Continuation (6 of
43	45.4	2.2	116277	1	X20249	Borrelia burgdorfe
44	45.2	2.2	2215	1	Q70882	Tomato spotted wil
45	45.2	2.2	2621	1	Q76109	Tomato spotted wil

## ALIGNMENTS

RESULT 1	
ID X02828	standard; cDNA; 2062 bp.
AC X02828	
DT 14-MAY-1999	(first entry)
DE W09905287 Seq ID 3.	
KW Plant; flavonoid 5-transglycosylation activity; 57G; variety; colour; ds.	
OS Verbena hybrida.	
EH Key	Location/Qualifiers
FT CDS	26..1411
FT /tag= a	
FT /product= "protein with flavonoid 5-transglycosylation	
FT activity"	
PN W09905287-A1.	
PD 04-FEB-1999.	
PF 16-JUL-1998; J03199.	
PR 25-JUL-1997; JP-200571.	
PA (SUNR) SUNTORY LTD.	
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,	
PI Yamazaki M.	
DR WPI: 99-142940/12.	
DR P-PSDB: W92949.	
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be	
PT used to transform plants for improvement of plant coloration	
PS Disclosure: Page 56-60; 89pp: Japanese.	
CC This invention describes the isolation of plant proteins which have	
CC flavonoid 5-transglycosylation (57G) activity. Such proteins can be used	
CC in the production of plant varieties with improved colour, for example in	
CC flowers harvested from them.	
CC Sequence 2062 bp; 608 A; 354 C; 508 G; 592 T;	

Query Match	100.0%; Score 2062; DB 1; Length 2062;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2062; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY 1	atttcccaaaaaataaaaaataatgacagagctcagctcctcttgccacattcc 60
DB 1	ATTTCACCAAAAAAATAAAAAAATGACAGAGCTCAGCTCTTGGCCACATTCC 60
QY 61	agcacagagacacataaatcccgccctcaattcgcccaagcgcttcgcaaatgcgacat 120
DB 61	AGCACAGGACACATTAATCCCGCCCTTAATTGCGCAAGCGTGTGCAAAATGCCACAT 120
QY 121	taagtacattcttcacagcgctcagcagcgctcagcagcgctcagcagcgcgctg 180
DB 121	TCAAGTCAATCTTCACACAGCGCTACATGCGCCCATGTCCAGAACCGCGCTGG 180
QY 181	cctcaacgggctcattcaattttgtcgctttccgacgggtatgcgaagggttcagcc 240
DB 181	CTCAAAAGGGCTCATCAATTTGTGTGCTTTCGACGGGTATGACGAGGGGTACAGCC 240
QY 241	cggagacgagtggaagactacatgctcgagagtgaaacagaggtatataagcccttgg 300
DB 241	CGGAGACGATGGGAAACACTTACATGTGCGAGATGAAAAACAGAGGTATATAAGCCTTGG 300

[illegible][illegible]

RESULT	2
X02826	
ID	X02826 standard; cDNA; 1506 BP.
AC	X02826;
DT	14-MAY-1999 (first entry)
DE	MO9905287 Seq ID 1.
KW	Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds
OS	Perilla frutescens.
EH	Key
FT	location/Qualifiers
FT	17..1399
FT	/*tag= a
FT	/product=
FT	activity"
PN	MO9905287-A1.
PD	04-FEB-1999.
PF	16-JUL-1998; J03199.
PR	23-JUL-1997; JP-200571.
PA	(SUNR ) SUNTORY LTD.
PI	Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI	Yamazaki M;
DR	WPI: 99-142940/12.
DR	P-FSDB; W92947.
PT	Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration

PS Disclosure: Page 49-52; 89pp; Japanese.  
 CC This invention describes the isolation of plant proteins which have  
 CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used  
 CC in the production of plant varieties with improved colour, for example in  
 CC the production of flower varieties with desired coloration and cut  
 CC flowers harvested from them.  
 SO Sequence 1506 BP; 364 A; 351 C; 449 G; 342 T;

Query Match 33.9%; Score 698.6; DB 1; Length 1506;  
 Best Local Similarity 71.2%; Pred. No. 1.6e-149;  
 Matches 1028; Conservative 0; Mismatches 364; Indels 51; Gaps 6;

```

OY 11 aaaaataaaanaaafagcagagctcagctccctctgtgcacactccacgaaggaga 70
DB 2 AAAATTTCCAAAATAATGTCGCGCGCGCGGTGCTGATGAACTTTCTGCGCAAGC 61
OY 71 cacataaaccgcgcctcactcagcagcgtctgcgaatgacagatcaatcaaca 130
DB 62 CACATTAATCCCGCCCTCAATTCGCCAAGACTCTTAAAGCCGCGACCTGACGTACA 121
OY 131 ttcttaccacgctctacgacgagcgcgcagatgtccagaacccgcgtgctcaacg- 189
DB 122 TTTTTCACGACGTTTATGATGCGCGCGCATGCGCAACAGCCTCCGCGCTCGCGGA 181
OY 190 -----gtccatcaatttctgtctgttccgaagcggtatgaacgaggttaccgcc 241
DB 182 AACCCACCGCGCGCTCGACTGCTGCGCTTCGCCAGGCTTCGACGAGCGGTGAAAGCC 241
OY 242 gagagcagtggaagaatcatatgtcgcgagatgaagaagaagatataaagccttggc 301
DB 242 TCGCGCGACGGGAAGCGCTACATGTCCAGATGAAGCCCGCGCTCCGAGGCTTAAAG- 300
OY 302 gatactctgagcccaataatgtcgaatacaaaaagcagaanaaatcgtctgtgtgac 361
DB 301 -----AAACCTCTCTCAAAACACGACGCTGACGCTTCGCTGCTAC 343
OY 362 tccacactctgtgacgagcgagcagagtgagtgagtgatctcctcgaagcgagccta 421
DB 344 TCCACCTCTTGTGATGGGCGCGGAGGTGCGCGCTGAGTCCAGGCTCCGCGCCCTT 403
OY 422 ctctgagatgagcagcagcagtggtgtgatatattacttatttcaatcaagcgcatagc 481
DB 404 CTCTGGGTGAGAGCCCGCACCGTGTGTGCAATATTTACTTCTTACTCAACGCTACGCA 463
OY 482 gacgaatcagatgagtggtgtgagatgatactgtcccgagagaccccaagtgctggc 541
DB 464 GACGAGATCGACGCGGTTCCGAGAAATGCA--GTCCTCGGCTTCACACCCCTGGAG 520
OY 542 cagcgtgattacgcgtcttctcttcttctcctcgaagcagatgagatctcgttcaatgat 601
DB 521 CAGCGCACTCTTCCGACCTTCTTCTGCTGCCGAGACCGAGAGTTCGGGTGATGATG 580
OY 602 aaggaagaatggaacttgaagaagtgaagaanaaacctaaagtgtgtgtaaacgctt 661
DB 501 AAGGAGAACTGGAACCTTTAGACGCTGAAGAAAGCGAAAGTTTGTGTAACACGTTT 640
OY 662 gatcgctgagagcctgagtcgctcaagcgcatgtgaaagtaagtaagatgacgcaatcg 721
DB 641 GATCGGTTGAGAGCCCGATGCACTGCGCTATGATAGTATGATGATGATGCGGATCGG 700
OY 722 cagtgatctcctcgaactctgtgagcagtaagaatcctcctcgagacaggtcttgcggga 781
DB 701 CCGTGTGATTTCCCTCGCTTCTTGTGACGCGGAGATTCCTCGAAACGTTTACGCGCGC 760
OY 782 gattgttcggaagaagtgtaagaagcagatgtcctcgaagtggttgaacgagatcct 841
DB 761 GATCTTTTCGAAAAA---TCGAGAGAAATTAATCTCGTGAATGATGATGACACGAAAGCC 817
OY 842 cgaatctcggtgttgaactgtcgttcggaagcctgtaataacagcagaagtcgcaaatg 901
DB 818 AAATCTTGAGTGTATGTGTCTTGTGGAGCGTTTGTAGAGTTTCAAAAGCACAAATG 877

```

```

OY 902 gaagatagcaagaaggcgtgttagatgttggaagccctttgtgggtgaagata 961
DB 878 GAGAGATTTGGAAAGGCGCTATTACCTCGGAAAGCCGTTTATGATGATACGAGAA 937
OY 962 aa-----cgaaaggaagaagtgattgataagatgtgacgtgaagtgagaa 1006
DB 938 CAGAAGATGACGACGCGGAAGAGAGAAAGAGATTGATGATGATGATGATGATGAA 997
OY 1007 cgaatggggaagaatgtatctgtgtgttctcaatgtggaagtcctgacatccctgtg 1066
DB 998 AAAATGGGGAATAATGTTTCGTGCTGCTCGCGCTGAGAGTTCGCGCACCTCGTTG 1057
OY 1067 gattgttcgtgacacactgcggtgtggaattcgaactcagaagatatacttgcgggtt 1126
DB 1058 GATTTTCGTGATCATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1117
OY 1127 cgaatgtgtgtcttccgagatgtgtcgaatgaagaagcagatgagcagatgagat 1186
DB 1118 CCGGTGAGGCGGCTCCGACAGTGTGATGATGATGATGATGATGATGATGATGATGAT 1177
OY 1187 gttgtgaggaagcgtgtgtgaagtgagcctaataggaagtgagcgtgtgtgtgtat 1246
DB 1178 GCGTGGGAGACAGGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1234
OY 1247 gaattagagatgtatgtgagagtgatgagtgagtgagtgagtgagtgagtgagtgag 1306
DB 1235 GAGATGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1294
OY 1307 gagagtgagcagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1366
DB 1295 GAAATGCAATTAATGGAAGCTTTGCGCAGAGAAAGCATGAGAGATGATGATGATGAT 1354
OY 1367 gtaacaacctcaagtgcttcttctgtagagtgagtgagtgagtgagtgagtgagtgag 1426
DB 1355 CTCAAGAAATCTCAAGCCTTCTTCTCATCAAGTTCACAGCTTATACAAATAAGCTT 1414
OY 1427 ccc 1429
DB 1415 TTC 1417

RESULT 3
X02827
ID X02827 standard; cDNA; 1474 BP.
AC X02827.
DE 14-MAY-1999 (first entry)
DT MO9905287 Seq ID 2.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Pellia frutescens.
FH key Location/Qualifiers
FT CDS 29..1360
FT /tag= a
FT /product= "protein with flavonoid 5-transglycosylation activity"
PN MO9905287-A1.
PD 04-FEB-1999.
PE 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y, Yamazaki M, WPI: 99-142940/12.
DR WPI: 99-142940/12.
DR P-PSDB: W92948.
PR Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration
PS Disclosure: Page 53-56; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut flowers harvested from them.
SO Sequence 1474 BP; 369 A; 348 C; 438 G; 319 T;

```

Query Match 33.0%; Score 680; DB 1; Length 1474;  
 Best Local Similarity 70.5%; Pred. No. 2,6e-145;  
 Matches 1019; Conservative 0; Mismatches 370; Indels 57; Gaps 6;

```

QY 9 aaaaaaaaaaataagacagagctcactctctgacacattccagcag 68
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 12 ACMAATTTCCACAAAATGTTCCGCCGCCGCTGCTAGCAAGTTTCCGCCCAAG 71
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 gaacataatccgcctcactcaatttcgcaagcgtctcgaattgcgaattcaagtc 128
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72 GCCACATTAATTCGCCCTCCCAATTCGCAAGAGACTCTCAAAAAGCGGACACAGCTCA 131
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 129 cacttcacacagcgtcactcagcgtcgcacgtccgaagacgcgcgtgctcaaac 188
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 132 CGTTTTTCAAGACGCTTTATGCAATGGCGCGCATGGCCAAACACAGCTCCGCCGCTGCC 191
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 189 g-----gctcaatcaatttctgcttctcgaaggtatgaacggttcaagc 239
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 192 GAATCCACACCGGCGCTGCACTGCTGCGCTTCCGACGCGCTAGACAGCGGCTGAAGC 251
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 240 cggagagatggaagaaactacatgctggaagtgaagaagcagaatataaagcttga 259
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 252 CCGCGCGCGACGGGAAGCGTACATGTCGAGATGAAGCCGCGCTCCGAGCGCTTAA 311
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 300 gcatatctctgacgcaataatcgtcgaataaagaagaacaaatcaagctgctgtgt 359
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 312 G-----AAACTCTCTTCTCAACAAGACGACGACGTCACCTTGTGCTGTCT 353
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 360 acctcacactcttgcataggcgagcgaagtgcgctgagctccatcccgagcgcg 419
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 354 ACTCCACACCTCTTGTGATGGCGCGAGGTGGCGTTTGTCCGACGTCACCGCCC 413
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 420 tacttgatctgacgacgaagtggttgatataattcttcttcttaacggctata 479
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 414 TTCTTGCGTGCAGCCGCCACCGCTGCTGCTATACCACTTCTACTTAAAGGCTACG 473
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 480 gcaacgaataatcgaatcggtctgacatactactcccggaagagatccagctctg 539
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 474 CAACAGAGATCGACGCGGTTCCAAATGAATCA---GCTCCCTGGCTTCCATTCCTGG 530
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 540 cccagcgtgattacgcgtcttccctctctccacgaatgaagatccgttcaactga 559
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 531 AGCAGCGCACTCTCGACCTTCTGCTGCTCGACGCGGAGAGATTCGGGTTGATGA 590
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 600 tgaagagaaattggaacttgaaggtgaagaaataaactaaggtctgtgtaaacgct 639
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 591 TGAAGGAGAACTGGAACCTTACACGCTGAAGAGAGCGGAAGATTAATGTTGAACACGT 650
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 660 ttgatcgctgagcgtgacgtcgaatgaatgaataagatgaatgacatgc 719
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 651 TTATGCGTTGGAGCCCGATGACATCGGCTATTGATAGTTATGATTTATGGGATCG 710
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 720 ggcgcgttgcctcgcacatctctgagcgttaagatccctcgaacggtcttcgcgc 779
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 711 GCGCGTGTGATTCCTCGCTGCTTGGACGCGGAAGATCCCTCGAAGACCTTACGCGCG 770
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 780 gaatcttgcgaagaaaggtcgaatgaacgaatgctcgaatggttgaacgaatc 839
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 771 GCGATCTTTTCGAAAA--TCGAGAGAGAAATACTCGGAGGTGATTAACCTCGAAGC 827
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 840 ctcgactctcggtgttactgcttgcgaagctcgttaataacgaacgtcgaac 899
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 828 CGAAATCTTGCGTGTATGCTGCTTGGAGCGCTTTGAGGTTTCCAAAAGCACAAA 887
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 900 tgaagagatagcaagagcgctgtagatgtggaagccgcttctgtgggtgtaaga- 958
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 888 TGAAGAGATTTGGAAAGGCTATTACCTGCGAAGGCGCTTTTATGATGATACGAG 947
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 959 -----gtaaagaaaggaagagatgataatgataatgataatgag 998
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 948 AACAGAAATGACGACGCGGAAGAAAGAAAGAAAGAGTTGATGGAG 1007
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 999 agtgaacagagtggaagaaattgtactctgtgtctcacttgaagatccctgacac 1058
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1008 AATTGAAAAAATGGGAAAAATGATGCTGCTGCTGCTGCAAGTTTGGAGTTTCCGCCACC 1067
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1059 cctcgttggagtgcttctgtgacacactgcggtggaattcgactctagaatatactt 1118
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1068 CTGCGTTGGGATGTTCTGTCGACGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1127
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1119 tcggagttccgagatggtggtcttccgcagtggtgtcatcaaggaagaaatgcaactga 1178
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1128 GCGGATTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1187
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1179 tgaagatgctgtggaagcgggtgtgagatggaactaagagagaggtgagctgtg 1238
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1188 TTAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1244
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1239 atggtatgaatgaagatgacttgaagaggttaltgaatggggagaaagaaatga 1298
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1245 ATGAGATGTCAGATGAGAAAGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1304
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1299 aacttagagagtgctgtgcaagtgaagatltggcaagaagaactatgaggaagatg 1358
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1305 TATGTGAGAGAAATGCTCAATGGAAGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1364
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1359 gatcttcagttacaacactcaaggtcttctgtatgagtgatgagttatcacaagcgt 1418
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1365 GATCTTCACTCAACAAATCTCAACGCTTCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1424
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1419 aatgag 1424
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1425 AAACAG 1430
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

#### RESULT 4

```

X02829
ID X02829 standard; cDNA; 1671 BP.
AC X02829;
DT 14-MAY-1999 (first entry)
DE MO9905287 Seq ID 4.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Tororia hybrida.
FH Key Location/Qualifiers
FT 45..1481 /tag= a
FT FT /product= "protein with flavonoid 5-transglycosylation activity"
FT FT
PI MO9905287-A1.
PI PN
PI PD 04-FEB-1999.
PI PE 16-JUL-1998; J03199.
PI PR 25-JUL-1997; JP-200571.
PI PA (SUNR.) SUNTORY LTD.
PI PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M.
PI DR MPI: 99-142940/12.
PI DR P-PSDB: W92950.
PI PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PI used to transform plants for improvement of plant coloration
PI PS Disclosure: Page 60-64; 89pp; Japanese.
PI CC This invention describes the isolation of plant proteins which have
PI CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
PI CC in the production of plant varieties with improved colour, for example in
PI CC flowers harvested from them.
SQ Sequence 1671 bp; 530 A; 344 C; 399 G; 396 T;

```

Query Match 23.5%; Score 484.4; DB 1; Length 1671;  
 Best Local Similarity 63.4%; Pred. No. 5.9e-101;  
 Matches 920; Conservative 2; Mismatches 453; Indels 75; Gaps 9;

```

QY 9 aaaaaaaaaaataagacagagctcactctctgacacattccagcag 68
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 31 AATTAATAAATAAATGTTTAAACAACGCTATTTCTACTACGAACATTCCTCCAGCAACAG 90
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

OY 69 gacacataaccgcccctcaattcgcgaagcgtctgcaaatgc-----cgacatcc 122
Db 91 GGCACATAAACCCCTTCTCGAGATTGCGCAAAAAGGCTCCTCAACACGGGATAGTGTGACC 150
OY 123 aagtcacattctcccaagcgtctacgcacatgagccgcatgctccaaagccgcgtgagct 182
Db 151 AGTGCATATCTTTCAGAGTGTATACGATGAGACGCAATG---CGCTTCGAACCGCATC 207
OY 183 caaacggcgctcaaatctgtctgcttccgaagcggtatgacgaagggtttacagcccg 242
Db 208 CGAGCAGCGAGATTCATTTCTGTGCACTKTYCAGATTTCTTACGATGATGCTTAAAGAAAG 267
OY 243 gagaagatggggaagaaactacatgctgggagatgaaagccgaaggtataaagccttgagcg 302
Db 268 GCGAGATGCGCAAAAACATACATGTGCGAGATGAGAAAGCGCGAAGCGCTTAAAGG 327
OY 303 atacctctgcagccaataatgctgcatacaaaaagcagcaca-----atcagct 350
Db 328 ACACCTTATTTAAGCTCAACGATGCGTGGCATGGGAAGTAATCTTACATCCGCTGAGCT 387
OY 351 tctgtgtgctactcccaactcttgcattggcgccgaaggtggcggtgagttccatctcc 410
Db 388 TTGTGTGTACTCTCATCTATTCTTCTGCGGAGCTGAAAGTGGCGGTGAATGCGAGCTGC 447
OY 411 ggaagcgctactctggtatgagcgaagcagcgtgtgtgatatatttactttatttca 470
Db 448 CGAGTGTCCCTTTTGTGGATTGAACGGCTACGCTTTCATGTGTATTTTACTTAACTTAA 507
OY 471 acggtatatacgcaagaatcgcgtgcgggtgctgcattgcattccgcggagagctcc 530
Db 508 ATGGGTATCCGATGATATATCGATGCGCGGTCAAGTCAATCCAGCTGCC---AATCTTC 564
OY 531 cagtgctggccagcggtgattacacgcttcttcccttccctccacgcatgagatctcc 590
Db 565 CCGACCTCTCCCAAGCAAGATCTCCCTCTTCTACTCTCCCTGAGACCCCGCGAGATGCC 624
OY 591 gttcaatgtagaagaaattggaaccttagaaggtgaaagaaaaaactaaggtcttgg 650
Db 625 GAACCTTAATGAAGAAAGTTTCGACACCTCTGACAAAGAACCGAAGGAAGGTTTGA 684
OY 651 tgaacagcttgcattgcgttgagccgtgcagcctgaagccattgaaagtacgagatga 710
Db 685 TAAACACGTTCCACGCAATTAGAAACCGAACAACCTCAAAAGCCATCGACAGTATGAATCTAA 744
OY 711 ttgcaactggcggttgcattcttccgcg---attctgtagcggttaagaatctcttcggaca 767
Db 745 TATCATGCGGCCCATTAATCCCATCATCATATTTCTACATGCGCAACGACCCCTCATCA 804
OY 768 ggtcttctggcgagagattgtctcgaaagaggtgcgaatgacgaatgctcctcgaaatgt 827
Db 805 GCAACAAATCTTACGATGAGACCTTTCAGAAAAGCCCATTAATAATTTACATGAGACTGTC 864
OY 828 tgaagcagaactcctgcattcgtgtgttgcgttctgcgttcgaagcctcgttaatacga 887
Db 865 TAAACTCAAAAACCCGATATGCGTGTTCGTTTACGTTTCGTTGCGGACCTTCTGAGGCTCC 924
OY 888 cgaagtcgcaaaatggaaagatatagcaagaaggtcgttgaattgttggagagccggttttgt 947
Db 925 CGAAACCCCAATATGGAAGAAATAGCAATATGGGCTTTCAGACACCAAAATGGCCAGTTCTGT 984
OY 948 ggtgtgttaagataaagaaagaaagagta-----ttgataagt 988
Db 985 GGGTATTAAGAAAGAAAGCAAGAGGGCGAGCAACAGACAGCAAGCAAGAAAGAAAGAAAC 1044
OY 989 tgcataagagagttga-----aaacagagtgaggagaatattgatatcttggttgt 1034
Db 1045 TCTGTAGCTTCTTGTATCTCTACAGCACTGAACGACTCGGGAAATATGTCATGATGCTGCT 1104
OY 1035 ctcaatttgaagatcctgcacatctccctgcttgagatgttctgcgttcgacactgcgggttga 1094
Db 1105 CACAATTTGATGTCTGTGACGCAATAGTCTGCTGGATGCTTCGTAAGCAGCATGCGGTTGGA 1164

```

```

OY 1095 attcgactctagaagatatacttctggggttccgaatggtgcttctccgacgttctcg 1154
Db 1165 ATTCTGTATTCAGAGACCTGGCTGTGTGTGTCGCCGTGTGTGCTTCTTCATGAGTTGG 1224
OY 1155 atcaagagcagaatcgaaagcgtatgagatgtgtgtggaagcagggtcgtgaagatgaag 1214
Db 1225 ATCAAGGAGACTAATGGAAGATGATGGAAGATGTGTGAGAGATGCTGTGAGAGTACAG 1284
OY 1215 ctaatgaagagaggtagcgtctgtatgctgataatgagagatgatatgagaggtta 1274
Db 1285 TGAATGAGGAAGCGCGCTTGTGTATGAGCGGTGAAGAGTGTGCTGAGAGCTTA 1344
OY 1275 tggatgggggaaagaaatagaagaaactagaagagagctcgcaagtgtgaagatgttg 1334
Db 1345 T-----AAAGATTCAGAGATTGAGAAAGCCCAATGATGTGAAGGCTTTGG 1392
OY 1335 caagaaagctatggagaga---tgatcttcaatcaaacctcaagctcttctcg 1391
Db 1393 CTAAAGAACCTATGATGATGAAGAACGTGATCATCATGAATCAATCTGAAGATTTTATTA 1452
OY 1392 atgaaggtgt 1401
Db 1453 CTAGCATTTAT 1462

RESULT 5
X02831
ID X02831 standard; cDNA; 2105 BP.
AC X02831;
DT 14-MAY-1999 (first entry)
DE W09905287 Seq ID 6.
KW Plant; Flavonoid 5-transglycosylation activity; 5tg; variety; colour; ds.
OS Petunia hybrida.
FH Key Location/Qualifiers
FT CDS 341..1747
FT /tag= a
FT /product= "protein with flavonoid 5-transglycosylation
FT activity"

W09905287-A1.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI; 99-142940/12.
DR P-PSDB; W92952.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PT used to transform plants for improvement of plant coloration
PS Disclosure: Page 67-71; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylation (5tg) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC flowers harvested from them.
CC Sequence 2105 BP; 669 A; 378 C; 429 G; 629 T;

Query Match 20.6%; Score 424.2; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred No. 2.8e-87;
Matches 856; Conservative 0; Mismatches 513; Indels 54; Gaps 7;

OY 14 aataaaaaaaatgacagcagcgtcctctcttgcacacatcccaagcagaggacac 73
Db 329 AAGCAAGAAATATATGTTGTCAGCTCATGTCACTTAAACAATTTCCACACAGGCCAT 388
OY 74 ataatacccgcccttaattgcgcaagcgtctcgcgaatagcgcgaacttaagtcaatcc 133
Db 389 ATTAATCCAGCACTTCAATTTGCAAGAAATCTGTGAATGCGCATTAAGAGCATTTT 448
OY 134 ttcaccagcgtctacgacgagcgccatg---ttcagaacgcgcgtggtccaacggg 190
Db 449 TCTACAGCATTTATGCTCCCAAGCCGTATGATGAATAATTCATTCTTATGACCAAAA 508

```

```

QY 191 ctacacatttctgtcttccacgagatagcagcggttacaccccgagacat 250
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 509 GGATTGAATTTTCATTCCATTTTCGATGGCTTTATGATGAGTTTGTATCAAAAGAC 568
QY 251 gggagaacacacatctcggagatgaaaagcagaggtataaagccttgagcactct 310
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 569 CCTGTATTTTACATGTCACAACTTCGTAAATGTTGAAAGTGAATGTCAAAAAATATAT 628
QY 311 gcagccataatgtcgaataaaaaagcagaataacgyltcggtgactccacctc 370
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 629 -----CTCACTTGCTGTGAAAATGACAGCCTTAACCTTGCTTACTTCATTTTC 682
QY 371 ttgcacatgagcgagcgaggtgagcgtgagttcacatcccgagcactcgtgatt 430
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 683 CTTCCTTGGGAGAGAGAGTACGACGTGAAGTTCACATCCCTTGCTCTTTGGAGT 742
QY 431 gagcagacacagtgctgagatacttaactttaattcaaggcctatagcgacgaatc 490
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 743 CAACCAACCAACAATATGACATATATTACTTCAACTTTCATGATATGAAAAAGCTATG 802
QY 491 gatgcggttcgagtgctatcactgccc-----gagagatcccgatgctggcc 541
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 803 GCTAATGAATCCAAATGATCCAAATGCTCACTTCCCGGCTTCCACTACTGAA 862
QY 542 cagcgtgattacgcgtctctcctcctcct-----ccacgatgagagattc 589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 863 ACTGGAATCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 922
QY 590 cgtcactgagtgagagagaaatggaacttagaaggtgagagaaaaacctaagctctg 649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 923 CTTCCACCATTCAAAGAATGATGATGACACATTAATGCTGAACCATCTCAATGATCTT 982
QY 650 gtgaaacgcttgctgctgagagcgtgagcgtgacatgataagtaagatg 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 983 GTGAATACATTTGATGATGATGAGCCTGAGCCTGAGCCTGAGCCTGAGCCTGAGCCTT 1042
QY 710 atgcaatcggcgctgattcctcctcgcgactctgagcgtgaaagcctcctcgagcag 769
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1043 TATGGAATGAGACCGTTCATCTTCTGCTTCTTCTGCTGGAATGACCTTTAGATGCT 1102
QY 770 tcttcggcgagagatctgctcgaagaaggtcgaaagcagatgctcctgaatgctg 829
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1103 TCATTTGCTGCTGATCTTTTCA-----AATTCAAATGACATATGAAATGCTTA 1153
QY 830 agcgcgaatcctcgtctcgtggttgaactcgttcgagaaactcgttaatacagc 889
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1154 AACTCAAGCCAAATTCATGCTGTTATATATCTTTGGAGCTTAATGAATTCATCT 1213
QY 890 aagtcgcaaatgagagatagcagaagcgctgttagatgtgagagcgcttctgtg 949
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1214 ATTAGCCAAATGAGAGATATCAAAAGCGTTGATAGACATAGAGAGCGCTTTTATAG 1273
QY 950 gtggtgaagatg-----aacgaagagagaaggtatgtataagttgcatgag 997
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1274 GTGATTAAGCAAAATGAAAAAGCAAGAGAGAGATATAAAAGCTTGCTGATTCGA 1333
QY 998 gagttgaaacgaagtgagaaatgtatctgtgtctcaatgtgaaagtcctgaagcat 1057
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1334 GAATTTGAAAAAATAGAAAAATAGTTCCATGCTGCTCACAACCTTGATTAACAT 1393
QY 1058 cccctgtgtgagatgttcgttacacactgcggtgagatcagatcagagatatact 1117
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1394 CCATCTTAGAGATCTTTGCTTCTCATTTGATGAGAAATTCAGCTTAGAGATTTAGCT 1453
QY 1118 ttcgggggttcgcagatgctcctcgcagtggttcgatacaagagcagaaatgcgaagctg 1177
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1454 TGTGAGATGCGCATGTTGCGCATTTCCATGACAGATCAATGACAAATGACCAACAA 1513
QY 1178 atgagagatgtgtgagagcaggtgtgagagtgagacataagagagggtaagcgtct 1237
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1514 GTTGAAGATGTGTGAAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1570

```

```

QY 1238 gatgtgatgaataatgagatgattgaggaggtatagatgagggagagaaagatag 1297
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1571 GAAAGTGAGGAAATCAAAAGCTGATTGAAATGCTAATGATGAGAGAGAAAGGGAA 1630
QY 1298 aaactagagagatgctgcgaagtggagaggtatggcaaaaagctatgagagagat 1357
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1631 GAATTTGAGAAAGAAATGCTAAGAAATGAAAGAAATGCTAGAGAACTGTCAAGAGAGT 1690
QY 1358 ggaactcgttaataaacctcaaggtctctcttgatgaaggtg 1400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1691 GGATCTTACACAAAGAAATTTAAAGCTTTATATGATGATG 1733

RESULT 6
T66166
ID T66166 standard; cDNA to mRNA; 1513 BP.
AC T66166;
DT 15-JUL-1997 (first entry)
DE UDP-glucose:thiohydroximate S-glucosyltransferase cDNA clone pGL9.
KW Glucosinolate; UDP-glucose:thiohydroximate S-glucosyltransferase;
   KM S-GT; transgenic plant; rapeseed oil; oilseed rape; canola; ds.
OS Brassica napus cv. Westar.
FH Location/Qualifiers
FT cds
FT 69..1469
FT 73
FT conflict
FT 73
FT /*tag- a
FT /*tag- b
FT /note- *C is T in clone pGL6-14, causing an amino
FT acid change from Ala-2 to Val-2"
FT 98
FT /*tag- c
FT /note- *G is A in clone pGL6-14; there is no amino
FT acid change"
FT 99..99
FT /*tag- d
FT /note- *codon AAG (Lys) is present between clone
FT pGL9 bases 98 and 99 in clone pGL6-14"
FT 195
FT /*tag- e
FT /note- *G is C in clones pGL6-14 and pGL3-22,
FT causing an amino acid change from Val-43
FT to Leu-43
FT 245
FT /*tag- f
FT /note- *T is C in clone pGL3-22; there is no amino
FT acid change"
FT 292
FT /*tag- g
FT /note- *T is C in clone pGL3-22, causing an amino
FT acid change from Leu-75 to Pro-75"
FT 323
FT /*tag- h
FT /note- *A is C in clone pGL3-22; there is no amino
FT acid change"
FT 331
FT /*tag- i
FT /note- *A is G in clone pGL3-22, causing an amino
FT acid change from Glu-88 to Gly-88"
FT 345
FT /*tag- j
FT /note- *A is C in clone pGL3-22, causing an amino
FT acid change from Asn-93 to His-93"
FT 354
FT /*tag- k
FT /note- *G is C in clone pGL3-22, causing an amino
FT acid change from Glu-96 to Gln-96"
FT 368
FT conflict
FT 368
FT /*tag- l
FT /note- *A is C in clone pGL3-22; there is no amino
FT acid change"
FT 377
FT /*tag- m
FT /note- *A is C in clone pGL3-22; there is no amino
FT acid change"
FT 377
FT conflict
FT 377

```

FT	acid change"
FT	398
FT	/*tag- n
FT	/note- "A is C in clone pGL3-22; there is no amino
FT	acid change"
FT	410
FT	/*tag- o
FT	/note- "A is T in clone pGL3-22; there is no amino
FT	acid change"
FT	458
FT	/*tag- p
FT	/note- "T is C in clone pGL3-22; there is no amino
FT	acid change"
FT	465
FT	/*tag- q
FT	/note- "A is C in clone pGL3-22, causing an amino
FT	acid change from Ile-133 to Leu-133"
FT	473
FT	/*tag- r
FT	/note- "T is C in clone pGL3-22; there is no amino
FT	acid change"
FT	512
FT	/*tag- s
FT	/note- "T is C in clone pGL3-22; there is no amino
FT	acid change"
FT	515
FT	/*tag- t
FT	/note- "A is T in clone pGL3-22; there is no amino
FT	acid change"
FT	526
FT	/*tag- u
FT	/note- "T is C in clone pGL3-22, causing an amino
FT	acid change from Ala-153 to Val-153"
FT	568
FT	/*tag- v
FT	/note- "C is T in clone pGL3-22, causing an amino
FT	acid change from Pro-167 to Leu-167"
FT	587
FT	/*tag- w
FT	/note- "A is G in clone pGL3-22; there is no amino
FT	acid change"
FT	623
FT	/*tag- x
FT	/note- "C is G in clone pGL3-22; there is no amino
FT	acid change"
FT	647
FT	/*tag- y
FT	/note- "G is C in clone pGL3-22; there is no amino
FT	acid change"
FT	678. . 679
FT	/*tag- z
FT	/note- "bases 678-679 are AT in clone pGL3-22,
FT	causing an amino acid change from Arg-204
FT	to Ile-204"
FT	685
FT	/*tag- aa
FT	/note- "C is T in clone pGL3-22; there is no amino
FT	acid change"
FT	714
FT	/*tag- ab
FT	/note- "A is G in clones pGL3-22 and pGL4-2,
FT	causing an amino acid change from Ser-216
FT	to Gly-216"
FT	759
FT	/*tag- ac
FT	/note- "G is A in clone pGL3-22, causing an amino
FT	acid change from Ala-232 to Thr-232"
FT	761
FT	/*tag- ad
FT	/note- "G is C in clone pGL3-22, causing an amino
FT	acid change from Ala-232 to Thr-232"
FT	766
FT	/*tag- ae
FT	conflict

FT	/note- "G is A in clones pGL3-22 and pGL4-2, causing an amino acid change from Arg-234 to Lys-234"	
FT	794	
FT	conflict	
FT	/tag- af	
FT	/note- "A is C in clone pGL3-22; there is no amino acid change"	
FT	809	
FT	conflict	
FT	/tag- ag	
FT	/note- "C is T in clone pGL3-22; there is no amino acid change"	
FT	814	
FT	conflict	
FT	/tag- ah	
FT	/note- "G is C in clone pGL3-22, causing an amino acid change from Gly-249 to Ala-249"	
FT	848	
FT	conflict	
FT	/tag- ai	
FT	/note- "C is T in clone pGL3-22; there is no amino acid change"	
FT	901	
FT	conflict	
FT	/tag- aj	
FT	/note- "C is T in clone pGL3-22; there is no amino acid change"	
FT	936	
FT	conflict	
FT	/tag- ak	
FT	/note- "G is A in clone pGL3-22, causing an amino acid change from Gly-290 to Arg-290"	
FT	959	
FT	conflict	
FT	/tag- al	
FT	/note- "C is A in clone pGL3-22; there is no amino acid change"	
FT	968	
FT	conflict	
FT	/tag- am	
FT	/note- "G is A in clone pGL3-22; there is no amino acid change"	
FT	973	
FT	conflict	
FT	/tag- an	
FT	/note- "A is C in clone pGL3-22, causing an amino acid change from Lys-302 to Thr-302"	
FT	1013	
FT	conflict	
FT	/tag- ao	
FT	/note- "C is T in clone pGL3-22; there is no amino acid change"	
FT	1117	
FT	conflict	
FT	/tag- ap	
FT	/note- "G is T in clones pGL2-7 and pGL2-25, causing an amino acid change from Val-350 to glu-350"	
FT	1121	
FT	conflict	
FT	/tag- aq	
FT	/note- "G is T in clone pGL2-7; there is no amino acid change"	
FT	1130	
FT	conflict	
FT	/tag- ar	
FT	/note- "C is T in clone pGL2-7; there is no amino acid change"	
FT	1161	
FT	conflict	
FT	/tag- as	
FT	/note- "T is C in clone pGL2-7; there is no amino acid change"	
FT	1169	
FT	conflict	
FT	/tag- at	
FT	/note- "A is G in clone pGL2-7; there is no amino acid change"	
FT	1193	
FT	conflict	
FT	/tag- au	
FT	/note- "T is G in clone pGL2-7; there is no amino acid change"	
FT	1202	
FT	conflict	
Query Match 8.7%; Score 178.4; DB 1; Length 1513;		
Best Local Similarity 53.0%; Pred. No. 1.4e-31;		
Matches 405; Conservative 0; Mismatches 356; Indels 3; Gaps 1;		

OY	629	gaaabaaabocctaaaggtcttggtgaagaagctttgaatgagcttgagagccctgaibgcctcaag	688
Db	667	CAACATGCTGATTGGCTGCTTCGTCACAAAGCTTGGAAGGTTTAAAGCACACAAAGCTTGTA	746
OY	689	gcacatgataagtaacgaatgatatgcacatgcggcgttgatccctccgcacatcttgagc	748
Db	747	GTTGAGAAATCAGAGGGGATGAGGGCAGCTTATCGAAGCTATGATACCACTGCTGTAAT	806
OY	749	ggtaagaatccttcggacaaagctcttcggcgaagaatttgtcgaagaagaaggtcgaatgac	808
Db	807	CTCAGCGCCCAATCAAAAGAGATMAAGGGCTACGGCTCAGAGCTGTATGAAGCGCGCTCTCG	866
OY	809	gagcatgcccgcgaatgggttgaagacaagaaccgcgaatcctcggtagtttaagcttcgtcc	868
Db	867	GAGCGATGTATGAGTGAGTGTTAGACACTAACCTTAGCAAGTGGTGATTTTGTGTGGTTT	926
OY	869	ggaaagcttcgttaataacgaagaaatcgcgaatggaagaaatagacaagaagagcgtatgaat	928
Db	927	GTTTCCTTTGGGATCCGCTTTGAGAAACACACTCGATGAGGTGGCAAAAGCGTTTACAGAA	986
OY	929	tgttggaagagccgcttcttggttggtgaagaagtacaagaagagaagatattgaat	988
Db	987	TCCAACCTTAACCTTCTTGCGGTATCAAAAGACGTCATATACGAAATTACACAGAGGG	1046
OY	989	tgcataggagaagctgtaaaacgaatgaggaaataatgtatcttggtgtctcaattgaagac	1048
Db	1047	TTTGCGAAGCTTACCAAAAGACAGAGCGCTTCTGTGTTCTTGGTGTAACCACTTGAGGTT	1106
OY	1049	ctgacgcacccctcgcttggaatgtcttcgttacacaactgcgggtggaatccgaatcgaag	1108
Db	1107	TTAGCTCATGTATGATAGTAAGTTGCTTTTGTACACACGCGGTTGGAACCTCGACGTTGGA	1166
OY	1109	agtaabctcttcggggtccgaaatgtagtgccttccgcgaatgtgtcgaatcaagaagacaat	1168
Db	1167	GGATGAGATTTGGGAGTTCGCAGTGGTGGTGCTCCATCAGCGAGTATCATGATGATGAT	1226
OY	1169	gcgaagacgaatgagaagatctgtgtgaagaacgggtgctgaagatgagaatgaagaaggt	1228
Db	1227	GCTAAGTTTGGGAGAGAGGTTTGGAGAGTTGGGTATGATACCGAAAGAGGAAAGCAACCTGGCGGA	1286
OY	1229	agcgatcgcttgaatgtgataaattagaagaatgataatgaagaagttatgatatgagaga	1288
Db	1287	GGAGTTTGCAAGACCAATGAGAGTGTGATGCTTTTGAGAGAGATGATGGA---AGCAGAG	1343
OY	1289	aagagtagaanaacttaagaagaagtgtcgtcgaatggaagaagatttgcgaagaagaagctatg	1348
Db	1344	AGTACTGTGGAGATTAGAGAGAGGTTTAAAGAGGAAAGATTGGCTGTGAAGCGGATG	1403
OY	1349	gaggaagaatggaatctcaatgataaacaactcaagatcttctctga	1392
Db	1404	AGTGAAGAGAGAGCTCTGATCGGAGAGATTAAAGATTTGTTGGA	1447

	RESULT	7
X02828/c		
ID	X02828	standard; cDNA; 2062 BP.
AC	X02828	
DT	14-MAY-1999	(first entry)
DE	W09905287	Seq ID 3.
KM	Plant; Flavonoid 5-transglycosylation activity; STG; variety; colour; ds	
OS	Verbena hybrida.	
FH	Key	Location/Qualifiers
FT	CDS	26..1411
FT		/*tag= a
FT		/product= "protein with flavonoid 5-transglycosylation
FT		activity"
PN	W09905287-A1.	
PD	04-FEB-1999.	
PE	16-JUL-1998;	J03199.
PR	25-JUL-1997;	JP-200571.
	(SUNR )	SUNTORY LTD.
PI	Gong Z, Sunitani T, Mizutani M, Saito K, Tanaka Y,	

PI Yamazaki M;  
DR PPI: 99-142940/12.  
DR P-PSDB: W92949.  
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
PS used to transform plants for improvement of plant coloration  
PS Disclosure: Page 56-60; 89pp; Japanese.  
CC This invention describes the isolation of plant proteins which have  
CC flavonoid 5-transglycosylolation (5TG) activity. Such proteins can be used  
CC in the production of plant varieties with improved colour, for example in  
CC the production of flower varieties with desired coloration and cut  
CC flowers harvested from them.  
SQ Sequence 2062 BP; 608 A; 354 C; 508 G; 592 T;

Query Match	Similarity	7.1%	Score 147	DB 147	Length 2062
Best Local	Similarity	63.1%	Pred. No. 2e-24		
Matches	330	Conservative	0	Mismatches 125	Indels 66
				Gaps	4
QY	1433	aggcaaaatctgcaaatctcattcgcgtgaagcttgaaactcttttgacttaatttgcga	1492		
DB	1921	AGGCGAAATTTGCAAAATTTGCTCAGGTAATTTACTACTCTTTCGGCTTTATTTATGTCTGA	1862		
QY	1493	gtttgtttttccaaaatt-attcttgaattctacaattgagctgaaattgactgattt	1551		
DB	1861	GATTGTGTTTTTAAATTTTCATCATATTAACCTTTTCATTGTGTCAAAAGTTAGTCCGATTTT	1802		
QY	1552	aac-----ttgaaaaataataaatctattgt	1578		
DB	1801	GGCTTAAAAAAATTCAAAAACTCGTTAGAAAATTCGGTGGAAAAATTTATGAACTCGTCCT	1742		
QY	1579	gagactcttcacaaataatcatctgattcttcttatctgctctgctcaaaaattccataca	1638		
DB	1741	CAGACACTTCATCAAGATGATCATTTGAATCCTTTATTTATTTTGGTCAAAATTTACTCTCTT	1682		
QY	1639	attgaaaaaataaatcttcaaaaatcgccaatcttgaaccagaagaagtataattga	1698		
DB	1681	CTTTGGTTCAAAATTTGGACGATTTTGAATTTTATTTTTCCAATGTATGAGAAATTTTGA	1622		
QY	1699	ccaaaataataaagaatcagaatgacttgacttgatgaagtctgcgaagagagctcatat	1758		
DB	1621	CCAGACAAATTAAGAAATCATGATGTTTTGTGAGGAAGCTCTCAACAAATGAATTTTATAT	1562		
QY	1759	ttctccaccgaattcttaacgagttcttgaattttttttagccaaatcsgnactaactt	1818		
DB	1561	TTTTTC-----AGTTAAATTCAGACTAAATTT	1535		
QY	1819	gtcaaaaatgaaaggtatatagtatgaaaatttaaaaaacaacaactcagacaataaaga	1878		
DB	1534	ACACTCAATGTAATAATTACAGAT-AAATTTTGAAAAAACAACCTCAACAATAATTTAAAC	1476		
QY	1879	cggaaagtgtataaattactctgcgaacgaatttgcgaatttcgctt	1921		
DB	1475	CAAAAAGTATTCACCTTAGAGATGAATTTTGGCAATTTTGGCT	1433		

RESULT	8
266173	
ID	T66173 standard; cDNA to mRNA; 566 BP.
AC	T66173;
DT	15-JUL-1997 (first entry)
DE	UDP-glucose:chlohydroximate S-glucosyltransferase clone pGL2-7.
KW	UDP-glucose:chlohydroximate S-glucosyltransferase;
OS	S-Glc; transgenic plant; rapeseed oil; oilseed rape; canola; ds.
PN	<i>Brassica napus</i> cv. Westar.
PD	EP-771878-A1.
PP	07-MAY-1997.
PR	31-OCT-1995: 402425.
PA	31-OCT-1995: EP-402425.
PI	(CANADA) NAT' RES COUNCIL CANADA.
PI	(PLBR2) PLANT GENETIC SYSTEMS NV.
PI	Grootwassink JMD, Hemmingen SM, Kolenovsky AD, Peferoen M;
DR	Reed DW, Underhill EW, Van Audenhove K;
	WPI: 97-247418/23.



PT Plants genetically transformed to interfere with  
PT UDP-glucose:chlohydroxamate 5-glucosyltransferase gene expression  
PT - useful for production or rapeseed oil with reduced glucosinolate  
PT content  
PS Example 2: Page 18-19; 35pp; English.  
CC Brassica napus UDP-glucose:thiohydroxamate 5-glucosyltransferase  
CC (5-GT) partial cDNA clones pGL2-7 (T66173) and pGL2-25 (T66114)  
CC were obt'd. by PCR-RACE (see also T66167-77) of B. napus cDNA.  
CC Analysis of the sequences revealed an open reading frame of about  
CC 470 bp, a 104 3' untranslated region and polyA+ tail for both  
CC clones. The amino acid sequence of the encoded protein revealed  
CC part of 5-GT peptide 2 (M09827), as expected because this sequence  
CC was used for PCR-cloning, and the complete 5-GT peptides 5, 6 (with  
CC one amino acid difference) and 7 (M09830-32). A full-length cDNA  
CC clone (T66166) for B. napus 5-GT (M09825) was subsequently obt'd.  
SQ Sequence 566 bp; 160 A; 54 C; 193 G; 159 T;

Query Match	6.9%	Score 143	DB 1	Length 566
Best Local Similarity	60.38%	Pred. No. 1e-23		
Matches 255, Conservative	0	Mismatches 165	Indels 3	Gaps 1

QY	970	agaagagagatattgataatgtcaltgcatlgaagagatgaaacagatgaggaaataatgtatctctg	10253
Db	13	AGCAACATTACACAAAGGTTTGTGAAAGCTAACAAAGACAGAGCGTTGCTTGTTCCTTG	72
QY	1030	gtgttcacaaattgaaagtcctcgaacacccctcgtctggagatgtttcgtacacactgcgg	10859
Db	73	GGTAAACACAGCTTAGGTTTAACTCATGATGATCATAGTGTGTTTTTGTGACACACTGGGG	132
QY	1090	gtgaaattcgactctagagaatatactcttcggggttccgatatgtgcttccgaaagt	1149
Db	133	TTTGAACTCCGACGCTGGAAAGGTTTGATTTGGGACGTTCCGATGCGTGGCGTGCCCCACTG	192
QY	1150	gtctcaacaaaggagacgaatcgcaagctcgatagagagatgtgtgaaagacgggtgtgaaggt	12093
Db	193	GAGTGATCATCATGATGATATCTCTAAGTTTGTGGAGAGAGGTTTGGACAGTTGGGTATAGAGCC	252
QY	1210	gagagcctaattagagagaggttagcgctcgtctgaatgtgtgaatlaagagatatgtattgaga	12659
Db	253	CAAGATGCAAGCTCGGGGAGGAGATTGGAAGACGATGAGTGTGTGAGGTGTTTGAAAGG	312
QY	1270	ggtcatatgagatggggagaaaaagatagaaccttgaagagagatgtctgccaagtgaaga	13289
Db	313	AGTATATGGA--AGGAACAGAGTAGTGTGCGAGATTACAGAACTTCTTAACAAATATGGAAGA	369
QY	1330	tttgacaagaagaagctcttgagagaaatgatctcaagtlaacaaacctcaagcttctct	13889
Db	370	TTTGGCTCTGTGAAGCGATGTAGTGAAGAGAGAAAGCTCTGATCGGACATTAAGAATTGTCT	429
QY	1390	tga	1392
Db	430	TGA	432

RESULT	9	
10106		
ID	10106	standard; cDNA to mRNA, 1731 BP.
AC	10106	
DT	27-OCT-1996	(first entry)
DE	Maize UDP-glucose 4ndol-3-yl:acetyl:glucosyl transferase sequence.	
KW	UDP-glucose-4ndol-3-yl:acetyl:glucosyltransferase; enzyme;	
KW	transgenic plant; EC-2.4.1.121; crop improvement; corn; cereal;	
KW	grass; IAGLU-transferase; ds.	
OS	Zea mays.	
FH	Key	location/qualifiers
FT	cds	57..1472
FT		/*tag= a
FT	misc_binding	858
FT		/*tag= b
PN	W09600291..A1.	/note= "start of UDP binding region"
PD	04-JAN-1996.	

PF 20-JUN-1995; 007820.  
 PR 24-JUN-1994; US-265427.  
 PA (RESE ) RESEARCH CORP TECHNOLOGIES INC.  
 PI Bandurski RS, Szczygłowski K, Szerszen JB;  
 DR WPJ; 96-068875/07.  
 DR P-PSDB: R89999.  
 PT UDP-glucose indol-3-yl:acetyl transferase DNA and protein  
 PT - used in sense or anti-sense orientation for the genetic control of  
 PT plant IAA levels and plant growth  
 PS Claim 2; Page 21-23; 41pp; English.  
 CC This nucleotide sequence encodes a UDP-glucose-indol-3-  
 CC yl:acetyltransferase (Indu-transferase) from maize (*Zea mays*).  
 CC Over-expression of this sequence in transgenic plants will result  
 CC in at least 1 of the following effects: inhibition of apical  
 CC dominance; inhibition of stem elongation; inhibition of cell  
 CC enlargement and increased numbers of stems per plant as compared  
 CC with a wild-type plant. The ability to control the expression of  
 CC this enzyme allows the control of free indoleacetic acid levels in  
 CC plants, thereby affecting plant growth rates. Antisense constructs  
 CC may be used to inhibit synthesis of this enzyme.  
 SO Sequence 1731 BP; 298 A; 528 C; 601 G; 304 T;

Query Match	6.88;	Score 141;	DB 1;	Length 1731;
Best Local Similarity	53.88;	Pred. No. 4.3e-23;		
Matches 316; Conservative	0;	Mismatches 265;	Indels 6;	Gaps 1;

QY	806	gaacgaagatctctcgaatagtttgagcaagaatctctgaatcttggttgtaagtttcg	865
Db	849	GAGGATGGGTCCACCAAGTGGCTTAGAACACCAACCCGACGGCTCGGTGGCCATCTTC	908
QY	866	ttcgaagcttcgfttaatacgaagaatctgcaatggaaagatatagcaagaaggtctga	925
Db	909	TTCCGCAACCTTCGGGCTCCCTGGGCAACGCCCAAGAGAGAGCTCCGGCGGCTCTC	968
QY	926	gatttgagagcgccgttttctgtgtgtlaagaatgaagaaagaaagagattatga	985
Db	969	GCCCGCCGCAAGCCGTTTCGTGGGTGTGAGGGCCAGAGAGACAGCATGTCCCGGC	1028
QY	986	aattgcatactgagga-----gttgaaacgagtgggaaatgtatctgtgtgtctcaa	1033
Db	1029	TATCTCCGTGGCCGAGGGAGCGGACACGGGCGCCGATGGTGTGCTTCCTGTGCCGAG	1088
QY	1040	ttggaagtcctgaagcaatccctcgtttggagtglttcgtgaacacatcgagggtgaaatcg	1098
Db	1089	CTGGACGTCGTGGGCGACCCGGCCGTGGCTGTCGTACCCACCACTGCGGTTGGAACTCC	1148
QY	1100	actctgaagatatactctcgtgggttcgacagatgtgccttcgcagcagtggtctgataa	1155
Db	1149	ACGCTGAGAGGGGCTCACCTTCGGGCTGCCCTATGATGGCATGGCGTGTGGACGGACCG	1208
QY	1160	ggagacgaatctgcaagactgatatgagaatgltgttgagagacggtgtgtgaagtgaagctaat	1215
Db	1209	CCGACCAACGCTCGGAACGTTCGACACTCGCTGGGCGGGCGTGGCGCGCGCCGCAAT	1268
QY	1220	gagagaggttagcgtctgtgatatgtgatagaatlaagaatgtcatgtgagaggtttatgata	1278
Db	1269	GCTGGCGGGCGGTTCCTTCGGGGGGAAGTGACCGGTGCGTGGCGCCCTCATATGAC	1328
QY	1280	ggggagagaaagagtagaagaacttagaagaagtgctgcgcagatgagagagattctgcaaga	1333
Db	1329	GGGGCGAGGGGGGCTGTGCTGCACGCAAGAGGGGGGGAATGAGAGGACAGAGGCTTGC	1388
QY	1340	aaagctatggagagaagatgatacttcaagtaaacacctaaggtctt	1386
Db	1389	GCCGGGTGGCACCCGGGTGGAGGTTCTGCACCGCAACCTGACGAGATT	1435

RESULT	10
ID	T66174 standard; cDNA to mRNA; 568 BP
T66174	
AC	T66174;
DT	15-JUL-1997 (first entry)

DE UDP-glucose:thiohydroximate S-glucosyltransferase clone pGL2-25.  
KM Glucosyltransferase; UDP-glucose:thiohydroximate S-glucosyltransferase;  
KW S-GT; transgenic plant; rapeseed oil; oilseed rape; canola; da.  
OS Brassica napus cv. Westar.  
EP-771878-A1.  
PD 07-MAY-1997.  
PF 31-OCT-1995; 402425.  
PR 31-OCT-1995; EP-402425.  
PA (CANADA) NAT RES COUNCIL CANADA.  
PL (PLANT GENETIC SYSTEMS NY.  
PI Groenwalsink JMD, Hemmingen SM, Kolenovsky AD, Referoen M;  
DR Reed DM, Underhill EW, Van Audenhove K;  
WPI: 97-247418/23.  
PT Plants genetically transformed to interfere with  
PT UDP-glucose:thiohydroximate S-glucosyltransferase gene expression  
PT - useful for production or rapeseed oil with reduced glucosinolate  
PT content  
PS Example 2; Page 19; 35pp; English.  
CC Brassica napus UDP-glucose:thiohydroximate S-glucosyltransferase  
CC (S-GT) partial cDNA clones pGL2-7 (T66173) and pGL2-25 (T66174)  
CC were obt. by PCR-RACE (see also T66167-72) of B. napus cDNA.  
CC Analysis of the sequences revealed an open reading frame of about  
CC 470 bp, a 104 3' untranslated region and polyA+ tail for both  
CC clones. The amino acid sequence of the encoded protein revealed  
CC part of S-GT peptide 2 (W09827), as expected because this sequence  
CC was used for PCR-cloning, and the complete S-GT peptides 5, 6 (with  
CC one amino acid difference) and 7 (W09830-32). A full-length cDNA  
CC clone (T66166) for B. napus S-GT (W09825) was subsequently obt.  
SQ Sequence 568 BP; 157 A; 55 C; 197 G; 159 T;

Query Match 6.5%; Score 135; DB 1; Length 568;

Best Local Similarity 59.1%; Pred. No. 6.7e-22;

Matches 250; Conservative 0; Mismatches 170; Indels 3; Gaps 1;

QY 970 agaaagatcgtatgaatgcatgaggaagtgaaacgagtgaggaaatgatactg 1029  
DB 13 AGCGAATCTACGAAAGCGTTTGTGGAAGCTACCAAGACAGACCGCTTGTCTTGG 72  
QY 1030 gttcttcattggaagctcgcagcaccctcgttgagatggttcgtgacaactcgg 1089  
DB 73 GTGTACCACTTGAGGCTTTTACCTCATGAAATGATAGTGTGCTTTTGACTCCTCGG 132  
QY 1090 gtggaatcagctcagatagatatacttcggggctgcagatggtgcttcctcagtg 1149  
DB 133 TTGGAATCTCAGCGTTGGAAGATTTGATTTGGAGTCCATGTTGCTGCTCAGTG 192  
QY 1150 gttcagatcaaggaagatcgaaagctgcatgagagatgctggaagcgggtgtgagatc 1209  
DB 193 GAGTGATCAGATGATGATGATCTTAAGTTTGTGGAAGAGAGTGGAGAGTGGTAAAGC 252  
QY 1210 gagaagcgaatgaggaagagtagcgtcgttgaatgagaaatgagagatgatatgagga 1269  
DB 253 GAAGGAGAGAGAGCTGGGGAGAGTGTGAAGAGGATGAGAGTGGTGGTGTGAGAGG 312  
QY 1270 ggtatgagatgaggaagaaagtagaagaaatagagagatgctgcaagtggaaagga 1329  
DB 313 AGTGATGCA---AGGAGAGAGTAGTGTGAGATTAGAGAGACTCTAAGAGTGAAGA 369  
QY 1330 ttgggaagaagaagctatgaggaagatgatacttcagtaacaactcaaggtcttct 1389  
DB 370 TTGGGCTGTGAAGCGCATGAGTGAAGAGAGAGCTGTGATCGAGCATTAATGATGTTGT 429  
QY 1390 tga 1392  
DB 430 GGA 432

RESULT 11

V17054

ID V17054 standard; cDNA; 1624 BP.

AC V17054;

DT 08-JUN-1998 (first entry)

DE Glucosyl transferase (Grase) encoding wound inducible gene (TW11).  
KM Glucosyl transferase; Grase; TW11; tomato; signalling pathway;  
KW salicylic acid; jasmonic acid; ethylene; wound inducible gene;  
KW plant defence protein; plant response; tobacco; rice; ss.  
OS Lycopersicon sp.  
FH Key  
FT Location/Qualifiers  
FT CDS  
FT 2..1410  
FT /\*tag- a  
FT /product- "glucosyl transferase"  
FT /note- "encodes Grase from amino acid position 5  
shown in W47172"

W09745546-A1.

PD 04-DEC-1997.

PF 30-MAY-1997; G01473.

PR 31-MAY-1996; GB-011420.

PA (UUYO-) UNIV YORK.

PI Bowles DJ, Calvert CM, O'Donnell PJ, Roberts MR;

DR WPI: 98-032653/03.

DR P-PSDB: W47172.

PT Tomato wound inducible (TW11) gene encoding glucosyl transferase -  
PT useful to develop products that alter signalling pathways in plants  
PT by altering of salicylic acid, jasmonic acid or ethylene

PS Claim 1; Fig 1; 52pp; English.

CC This wound inducible gene (TW11) isolated from wounded tomatoes encodes  
CC a glucosyl transferase (Grase) protein. This TW11 gene can be used to  
CC identify homologue Grase encoding genes isolated from tobacco and rice.

CC A microbial host can be transfected or transformed with a vector  
CC containing the Grase encoding nucleic acids. The products can be used to  
CC interfere with Grase and therefore alter signalling pathways in plants,  
CC specifically tobacco, rice or tomato plants by altering levels of  
CC salicylic acid, jasmonic acid or ethylene. This can induce the  
CC production of plant defence proteins such as pathogenesis-related (PR)  
CC (plant growth, reproduction and senescence) and improve plant response to  
CC pathogens.

CC Sequence 1624 BP; 545 A; 252 C; 372 G; 455 T;

SQ

Query Match 4.7%; Score 96.4; DB 1; Length 1624;

Best Local Similarity 50.7%; Pred. No. 5.4e-13;

Matches 291; Conservative 0; Mismatches 271; Indels 12; Gaps 2;

QY 804 atgacgaagattgctcgaatggttgagcagaatcctcgaatcctcggtggttacgltt 863  
DB 780 ATGAAACCGCGTCTGTAATAGCGTTGATTCGAAGAAATCAAGTTCCATTGATGTTT 839  
QY 864 cgttcggaagctcgttaatacagcagaagtcgcaatggaagagatgcaagagctgtc 923  
DB 840 GTTTGGAAAGTACAGAGATTTCCTACAGCAGATGCAAGAACTTGCTATGGGCTAG 899  
QY 924 tagatctgtggaagcggcttttctggtgtgtaagagtaaacgaaggaagaggtatgga 983  
DB 900 AAGCCTCTGGACAAGATTTCATTGGGTTATCAGAACAGGAATGAAGATTGGCTCCAG 959  
QY 984 taagttcagatgaggaagttgaaacgaatggaagaaatgtaattcgttcttct---caat 1040  
DB 960 AAGGATTTCGAGAAAGAAAGAAAGAAAGAGTTTAATCATAGAGATGGGACACCCAAA 1019  
QY 1041 tggaaagtcctgaagcaccctcgttggatggttctgagcaactcgggttggaattcga 1100  
DB 1020 GTGTGATTTCTTGAACAGAAAGCTATTGAGCTTTTGTACTCATTTGATGAGTAATCGA 1079  
QY 1101 cttcagaagatatacttcctgggttccgagtggtggtcttctcgaatggttcgataaag 1160  
DB 1080 CACTGGAAGGAATATCAGCAAGCGGTACCAATGTGAGATGCGCCAGTATTTGCCGAAGT 1139  
QY 1161 ggaagaaatgcaagctgagagagatgctggaagcgggtgagagtgagagataatg 1220  
DB 1140 TTTTCAATGAGAAAGTTGCTGATCAGTAAATGAGAAATGGAAGCTGGTGTGTTCTAAGC 1199  
QY 1221 aggaaggttagcgtcgttgatgagtgataaattagagagatgatgagaggttagatg 1280  
DB 1200 AATGGAAGAGAACAGCTAAGTGAAGAGAGTGAAGAGAGAAAGCAATAGCAAGGCGATTAAGA 1259



Oy	864	cgctccgaagagcttcgttaataaagaaagtcgcaaatcggaagagctgaagagagcgcgtc	923
Db	803	gTTTTGGCACTGTAAACAATTAACCCCTTAACGAATTTGGGTGATTAACAACACCTTAG	862
Oy	924	taagatctgaggagccgctttcttgaggctgaagataagatacgaagaggaagattatga	983
Db	863	AAACTAATAAAGAGCCCTTTATTTGGTCATTAAGAATAATATGCTGCAAGAAATTTGCCTA	922
Oy	984	taagctgcacatgagagagcttgaaaaagagctggggaagaaattgatacttgctgctcaattg	1043
Db	923	AAGCATTTCTTGAAGAAAGCAAGAACGACTTTGGGCAAAATATGTTCTTGGGCACCCCAATYGG	982
Oy	1044	agagccgcgaagcattccctctgttgagatgctttcttgacaaactgaggggctggaattccgctc	1103
Db	983	AAATCTTTGGCACATTAAGCTCTGTGGGTGTTTGTGAACGATTTGGGATGGAAATTCGATTT	1042
Oy	1104	taagagataatactcttcgcgggtccgagatgctgacttccgcagctgctcatcaagaga	1163
Db	1043	TGGAAGGCAATTTATTTGGTGTCCATATATATATAGGCGCTTTTGTGTGACCAAAAT	1102
Oy	1164	cgaaatcgaaagcctgcataggagagatgcttgagaagacgggctgagagctgaagctaatgag	1223
Db	1103	TGAATTAAGTAATAATGCTGGAAGAGCTGTGGGAAATGCTTTGCAAAAT-----TG	1150
Oy	1224	agggtaagcgtcgtctgagatgctgataataggagataatgataagagcttaagatggg	1283
Db	1151	AAGTGGGCAATTTTACTTAAGAATGGAATTAATATATATAGTCTTTGATATCTTTCTTCANTAG-	1209
Oy	1284	gagaaaaagataggaanaacttagagagagctgctgcagctgagagagatttggcaagaanaag	1343
Db	1210	--GAAAAAGGGAAGATATTTAAAGGAAATGTTGAAGGCGCTAATAAGAAACATTTAGAG	1267
Oy	1344	ctatagga--ggaagatgagatcttcagtttaacaacctcaagagcttctctgtagagtgct	1401
Db	1268	CTGGCAACCAAGATCATGGAAGTCAACAGAAAATTTCAAGGTTCTATGTTGAGCTAGTTAA	1327
Oy	1402	aggtatctaaagaagctaaatgagagctcccaatgagcaaaatgcaaatctcatctcgttaag	1461
Db	1328	ATGTGCACAGACCACTTAATATATTTAAATGTGACCGGTGACTTTGCATCTATAGTATCA	1387
Oy	1462	ttgataacttttggcttaaatcttgctcgaagcttgcttttcaaaattcatctctgtaatt	1521
Db	1388	ATATAT--CTTTGTGATGGGTTCTTTCTCGAAATTTGCAATTTGACAGGCTTAGCATATTA	1446
Oy	1522	ttacaatgagatgtaaatatttagcttgatttaactctggaataataataaatcaatgctgag	1581
Db	1447	TAAATGGTATCGGCGACTTTGTGTATTTTAACTGTGAACCTTTGCTCGAGTATGTATA	1506
Oy	1582	actcttcatacaaatcatcgtatcttcattatgcttccttgctcaaaaattctcatcaatt	1641
Db	1507	TAAATCCTCCTGTGACATATTTTAACTGCAATTTATTTGACTACTTCTTTTAAAAAA	1566
Oy	1642	ggaaaaaaataaa	1653
Db	1567	AAAAAAAAAAAA	1578
RESULT	14		
ID	T68693		
AC	T68693		
DT	05-JAN-1998	(first entry)	
DE	Strawberry UDP-glucuronosyl transferase cDNA.		
KW	Strawberry; UDP-glucuronosyl transferase; ERTID; fruit; ripening;		
OS	Fragaria ananassa Duch. cv. Brighton.		
PN	MO9721816-A1.		
PD	19-JUN-1997.		
PF	12-DEC-1996.		
PR	13-DEC-1995; GB-025459.		
PA	(ZENE ) ZENECA LTD.		
PI	Manning K;		

DR MP1: 97-332787/30.  
 PR Vectors encoding new enzymes for regulating ripening of fruit - used  
 PT particularly to improve storage properties, processing  
 PT characteristics etc, especially in strawberries  
 PS Claim 1: Page 23-24: 43pp: English.  
 CC This cDNA sequence comprises a UDP-glucuronosyl transferase (ERT  
 CC 1b) cDNA clone isolated from a ripe strawberry fruit cDNA library  
 CC by differential screening. 9 Ripening-enhanced clones (see T68687-  
 CC 95) clones were isolated from the library, and are deposited at  
 CC NCIMB 40693. Any of these 9 sequences can be used as a regulatory  
 CC sequence in claimed vectors for genetic transformation of plant  
 CC cells to regulate fruit ripening. The vectors also comprise a  
 CC promoter and transcription termination sequence. Also claimed are  
 CC plants (especially strawberry) and propagating materials containing  
 CC these vectors. Stable integration of the regulatory sequences, or  
 CC their complements or genomic equivalents, into a plant is used to  
 CC modulate fruit ripening by overexpression or downregulation of an  
 CC endogenous plant gene. Slowing ripening of fruit improves  
 CC resistance to damage during harvesting, packaging and transport.  
 CC It also extends shelf life, improves storage properties, processing  
 CC characteristics, flavour and aroma, and modifies colour and  
 CC Sequence 432 BP; 101 A; 84 C; 136 G; 94 T;  
 SQ

```

Query Match      4.0%; Score 83.4; DB 1; Length 432;
Best Local Similarity 56.4%; Pred. No. 3.1e-10;
Matches 150; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```

QY	993	tgaggagagctgaaacacggtgagggaataatgtatcttggtgtcttcataatttggaagctccga	1052
	1111		1111
Db	92	TGGAAAGGACGGAGACAGACAGAGCAAGGCAATGTTGCAATGGAGTCCACAGAGAAATTTGG	151
QY	1053	cgcatccctcgtcttgysga tgcctctgcacacactcgcgcgtgcgaattcgcactctagaagta	1112
	1111		1111
Db	152	AGCATCTTTCAGACGGCTTGGCTTTGTGACTCATTTGGCGGTGGAAATCAACCATGGAGTAC	211
QY	1113	tatctcttcggggtctccgaatggttggtcctcttcgcgaagtgcgttcataagagagaaatgcga	1172
	1111		1111
Db	212	TCACCTCAGGAATCCCGCTGATGTGCAATCCCAAAATGGGGTATCCAAAGTACCGCGCA	271
QY	1173	agctgatagaaagatctgttgagagacggtctgcagctgaagactaaatgagagaggtatagc	1232
	1111		1111
Db	272	AGTATTGTGTCGACGACGACTTTAAAGGTGGGAATGAAGATGTGCCGNNGAGAGGNCAGAAC	331
QY	1233	tcgctgatgctgatgataaattagagaga	1258
	1111		1111
Db	332	GGGTNNATCCCTAGGGAAGAGGTGA	357

RESULT	15	
V23108		
ID	V23108	standard; cDNA to mRNA; 1597 BP.
AC	V23108	
DT	06-AUG-1998	(first entry)
DE	UDP-glucose:flavonoid 3,5-O-glucosyl transferase gene.	
KM	UDP-glucose:flavonoid 3,5-O-glucosyl transferase; gentian; rose plant	
KW	blue flower; ds.	
OS	Unidentified.	
PH	Key	Location/Qualifiers
FT	CDS	39..1460
FT		/*tag= a
FN	J10113184-A.	
PD	06-MAY-1998.	
PF	15-OCT-1986; 272253.	
PR	15-OCT-1996; JP-272253.	
PA	(TOLG ) TOKYO GAS CO LTD.	
DR	WPI: 98-315473/28.	
DR	P-RSDB; W56451.	
PT	Gentian (flavonoid glucosyl transferase gene - used to produce	
PT	recombinant rose plants with blue petal(s)	
CC	Claim 3; Pages 7-8; 10pp; Japanese.	
CC	The present sequence encodes a UDP-glucose:flavonoid 3,5-O-glucosyl	

CC transferase enzyme. The nucleic acid sequence was amplified from  
CC RNA which had been isolated from gentian (sic) petals. The gene is  
used to transform rose plants to produce blue flowers. The gene is  
SQ Sequence 1597 BP; 483 A; 272 C; 368 G; 474 T;

Query Match 3.2%; Score 66; DB 1; Length 1597;

Best Local Similarity 48.3%; Pred. No. 4.2e-06;  
Matches 258; Conservative 0; Mismatches 260; Indels 16; Gaps 2;

```
QY 1028 tgggttcacatggaagctcgcgacatccctcgttggagatgttcgtgacacactgc 1087
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1056 TGGGCCCCACAGACGGCGGTGTGTCTCAGCTGTGGAGGTTTGTGTGCGATGT 1115

QY 1088 ggggtggaattcgactctagagagatatactctcggggtccgagtggtcttccgag 1147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1116 GGATGGAATTCAACACTGGAAGCATTTGGTGGTGTCCAGTGCAACCTGGCCCTCG 1175

QY 1148 tgggttcacatggaagctcgcgacatccctcgttggagatgttcgtgacacactgc 1207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1176 TTTGCTGAGCAGCATGATGATGATTTGAATTTGGTGAAGAAATTTGGCTTGGCTGGAG 1235

QY 1208 gtgagagctaaatgagagagtagcgtctgagtgagtgaaattagagatgatatgag 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1236 GTTAAGATGATTTATAAAGATTACAAAGATCTGACGGGAGCAATAATGTGAGAGCT 1295

QY 1268 gaagttatgagtgaggagaaagagtagaaacttagagagagtgctggcaagtggag 1327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1296 GATGTTATAGAGAGAAATCAAGATATTGATGATCTCGAAATGGAATCAGAAAGAAA 1355

QY 1328 gatttggcaagaaaa-----gctatgaggaagatgagatcttcaagtaac 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1356 GTGAAGGAATGAAAGAAAGACAGCGCTGGCTGTTGAGAGAGAGATCATCTTCTGCG 1415

QY 1373 aacctcaaggtcttcttgatgaggttgtaggtatcctaaagacg-taaatgaggtccca 1431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1416 TCCCTCAAGATTTTATATGATGTGATCAAAAGACATTCATGATTAAGCTTCCAAA 1475

QY 1432 taggcaaaattgcaaatctcgttaagtgaatacttttggcttbaatttgttcg 1491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1476 TTGGTTCAATAGATTTTATGGGCAAAACGTGAAGTGTCTTTGCAATTAGAGCAAAATGC 1535

QY 1492 agttgttttccaaaattatcttgtaattttacattgagtgtaattagct 1545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1536 TCTGATTTAATGTAAAGTTTATTTTATTTTATTTATCTGAATTAAGCTT 1589
```

Search completed: August 1, 2000, 18:17:46  
Job time: 1987 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:16:29 ; Search time 152.01 Seconds  
(without alignments)  
1865.931 Million cell updates/sec

Title: US-09-147-955-5

Perfect score: 2062  
Sequence: 1 atttaccacaaaataaaaa.....taaaaaaaaaaaaaaaaaa 2062

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/PCrus.COMB.seq: \*  
7: /cgn2\_6/ptodata/2/ina/Backfilest.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	141	6.8	1731	4	US-08-466-583-1
2	141	6.8	1731	6	PCT-US95-07820-1
3	96.4	4.7	1627	5	US-09-106-464-1
4	68.2	3.3	7218	1	US-08-232-463-14
5	59.6	2.9	1607	1	US-08-797-226-1
6	52.8	2.6	1669	3	US-08-522-421-1
7	48.4	2.3	1511	1	US-07-991-867B-8
8	48.4	2.3	1511	2	US-08-107-755A-8
9	48.4	2.3	1511	4	US-08-544-332-8
10	48	2.3	405	4	US-08-975-316-32
11	47	2.3	6768	2	US-08-107-755A-1
12	47	2.3	8457	1	US-07-991-867B-1
13	47	2.3	8457	4	US-08-544-332-1
14	45.6	2.2	4821	5	US-08-913-374-1
15	45.4	2.2	5852	1	US-07-867-106-2
16	45.4	2.2	19124	4	US-08-487-826B-13
17	45.2	2.2	2621	4	US-08-553-619B-8
18	44.6	2.2	1511	1	US-07-991-867B-8
19	44.6	2.2	1511	2	US-08-107-755A-8
20	44.6	2.2	1511	4	US-08-544-332-8
21	43.8	2.1	1972	2	US-08-463-048-1
22	43.8	2.1	1972	2	US-08-463-229-1
23	43.8	2.1	1972	3	US-08-302-891-1
24	43.8	2.1	19124	4	US-08-487-826B-13
25	43.4	2.1	665	4	US-08-883-795A-36
26	42.6	2.1	1738	3	US-08-379-482A-2

c 27	42.4	2.1	3440	2	US-08-471-791-27	Sequence 27, Appl
c 28	42.4	2.1	3440	6	PCT-US91-01746-27	Sequence 27, Appl
c 29	42.2	2.0	4098	4	US-08-605-106-4	Sequence 4, Appl
c 30	42	2.0	660	1	US-07-991-867B-32	Sequence 32, Appl
c 31	42	2.0	660	2	US-08-107-755A-32	Sequence 32, Appl
c 32	42	2.0	660	4	US-08-544-332-32	Sequence 32, Appl
c 33	41.6	2.0	731	2	US-08-451-405A-2	Sequence 20, Appl
c 34	41.6	2.0	246240	3	US-08-724-394A-20	Sequence 21, Appl
c 35	41.6	2.0	246240	3	US-08-724-394A-21	Sequence 22, Appl
c 36	41.6	2.0	246240	5	US-08-724-394A-22	Sequence 14, Appl
c 37	40.4	1.9	7721	5	US-07-867-106-2	Sequence 17, Appl
c 38	40	1.9	5852	1	US-07-867-106-2	Sequence 1, Appl
c 39	39.8	1.9	4379	2	US-08-592-214A-17	Sequence 1, Appl
c 40	39.8	1.9	6768	2	US-08-107-755A-1	Sequence 1, Appl
c 41	39.8	1.9	8457	4	US-07-991-867B-1	Sequence 1, Appl
c 42	39.8	1.9	8457	1	US-08-544-332-1	Sequence 9, Appl
c 43	39.6	1.9	1172	1	US-07-945-288-9	Sequence 9, Appl
c 44	39.6	1.9	1172	1	US-08-462-831-9	Sequence 9, Appl
c 45	39.6	1.9	1172	2	US-08-461-809-9	Sequence 9, Appl

## ALIGNMENTS

RESULT 1  
US-08-466-583-1  
; Sequence 1, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szczygiowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,583  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Feider, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELE: 49617824  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1731 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 57..1472  
; US-08-466-583-1





```

: GENERAL INFORMATION:
: APPLICANT: Steffens, John C.
: APPLICANT: Changas, Gurdsey S.
: APPLICANT: Kval, Jian-Ping
: APPLICANT: Eannetta, Nancy
: TITLE OF INVENTION: Chain Length Specific UDP-Glc:Fatly Acid
: TITLE OF INVENTION: Glucosyltransferases
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Jones, Tullar & Cooper, P.C.
: STREET: P.O. Box 2266 Eads Station
: CITY: Arlington
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/106,464
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/055,554
: FILING DATE: 13-AUG-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Spector, Eric S.
: REGISTRATION NUMBER: 22495
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-415-1500
: TELEFAX: 703-415-1508
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1627 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1413
: US-09-106-464-1

Query Match          4.7%; Score 96.4; DB 5; Length 1627;
Best Local Similarity 50.7%; Pred. No. 1.2e-15;
Matches 291; Conservative 0; Mismatches 271; Indels 12; Gaps 2;

QY 804 atgacgacgattgctcgaatggttgagacgaatcctcgatcctcgglttcgltt 863
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 794 ATGAACACGCGCTGTTGAAATGGCTTGATTCGAGAAATCAAGTTCATTGTTT 853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 864 cgttcggaagcttggtatatacgaagaagtcgcaatggaagagatagaagggcgt 923
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 854 GTTTTGGAGTACACGACGATTTCTACACGACAGATGCAAAAGCTTACTAATG 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 924 tagattgcyggaagcgcttttctggtgcytaagaagtaaacgaagaagaggtatga 983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 914 AAGCCTCTGGACAAAGATTTCATTGTTGTTATTCAGAAACAGGGAATGAATGGCTCCAG 973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 984 taagtgcattggaagagittgaacacgaagtcgggaaattgtatcctgtgtctc---caat 1040
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 974 AAGGATTCGAGGAAAGAAAGAAAGAAAGGTTTAAATCATTAAGAGATGGGACCCCAAG 1033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1041 tggaaagtcctgacgacatccctcgttggtgagtggttcgtacacacacgcgggtggaattgga 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1034 TGCTGATTCCTTGATTCACGACGCTATTGGAGCTTTTGTACTCATTTGTGATGGAAGACTGGA 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1101 ctctagagatatatcttcggggttcggaatggtgcttccttcgcaagtcggttcgaatcaag 1160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1094 CACTGGAAGGATATACACGAGGGGTCACCAATGTTACATGGCCAGTATTTCGGGGAACAGT 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

QY 1161 ggaacgaatgcgaagctgtatggaagatgtgtggaagcgggtgtgagagtgaagactaatg 1220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1154 TTTTCAATGACAGAGTTGTTGACTGAGTAAATGAGAGTGGACCTGGTGTGTTGTTAAGC 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1221 aggaagtagcgctgtgatggtatgaattagaagatgtatctaggaaggtatgagtg 1280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1214 AATGGAAGGAAACAGCTAGTGAAGAGTGAAGGAAAGGAAAGCAATAGCAAGCGATMAAGA 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1281 gggga-----gaaagagtagaactagaagagtgctgccaagtggaaagatt 1331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1274 GAGTAAATGCGCAGTGAAGAAACAGAGGATTCAGAAAGCAGACAAAGGATCAAGAA 1333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1332 tggcagaagaagctatggaaggaagatggaatcttc 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1334 TGGCAAGAGAGCTATTGTAAGAGAGGATCATTC 1367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-08-232-463-14/C
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHIEFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)883-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpt-F18
: US-08-232-463-14

Query Match          3.3%; Score 68.2; DB 1; Length 7218;
Best Local Similarity 3.4%; Pred. No. 4.5e-08;
Matches 10; Conservative 192; Mismatches 95; Indels 0; Gaps 0;

QY 1158 aaggacgaatgcgaagctgtaggaagatgtgtggaagacgggtgtgagagtcgaagacta 1217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```









FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/713,000  
FILING DATE: September 11, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: SLEATH, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000/1003C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 405 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-975-316-32

Query Match 2.3%; Score 48; DB 4; Length 405;  
Best Local Similarity 52.2%; Pred. No. 0.0018;  
Matches 167; Conservative 0; Mismatches 135; Indels 18; Gaps 2;

OY 799 gtgcgaatgcgaagatgctgtaagtgatgagcagaatccctcgatcttcggtgtta 858  
DB 86 GGCAGAGAGACGGAGTTCCTTCCTGCTGGACAGACGCCCTTACTCACTGCTTA 145  
OY 859 cgttcgctcggaagctcgtaatacagaagtcgcaaatggaagagatgaagaag 918  
DB 146 TGTGAGCTGTGGAGATCGCCTGTGTAACGAGTCGGAATTTCCGAATTAAGCTTAGG 205  
OY 919 gctgtgagatgtggagagcgcttctgtggtgtgaaga-----gtaaa 963  
DB 206 TTTAGCCGATAGACACAGCATCTTGCTGCTGACCCGGCTCAGTGAAGCGCTC 265  
OY 964 cgaagagagaagagatgtaagttgc--atggaagagtgaaagagtgaggaat 1020  
DB 266 GGAACCTTGAAGAAATTTGCCCGTCTTCTTGGAGCATTACAGAGAGGGGAAGAT 325  
OY 1021 tgaatcgtgtctcaatctgaagtcctgacgacacccctgttgagatgttcgtgac 1080  
DB 326 TGTGAATGAGGCGCTCAACATGAAGTGTGCTCATCGGCTGTGAGACGTTTGAGAC 385  
OY 1081 acacgtcggtggaattcga 1100  
DB 386 TCACATGGATGGAAGCTCCA 405

RESULT 11  
US-08-107-755A-1  
Sequence 1, Application US/08107755A  
Patent No. 5721352  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/107,755A  
FILING DATE: 19-AUG-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,658  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF114.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 372-5800  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6768 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Amsacta moorei entomopoxvirus

Query Match 2.3%; Score 47; DB 2; Length 6768;  
Best Local Similarity 46.2%; Pred. No. 0.012;  
Matches 270; Conservative 0; Mismatches 305; Indels 10; Gaps 3;

OY 1477 cttaaatgttcgagttgttttcaaatattatctgttaatttaacttgagtgtaa 1536  
DB 1018 CTCCTGTTTGGATTAATTCGATTTTGAAGAACCTTTCATATATATGAATGC 1077  
OY 1537 attagctgatttaactggaataataataatctgttgagactctcaataat 1596  
DB 1078 CTTATATATTTTGAAGTAAATAAGTATATTTATTTTATCTGTAAATAT 1137  
OY 1597 catctgattccttcatctgtctgtaaaatctcaatctgaatggaataataat 1656  
DB 1138 TTTTATATAGCAAAATVAGAAAAATTTCTTAAATATATTTTCAAAATTAATAT 1197  
OY 1657 caaatcgtccaattctgaaacgaagaagataatctgaccacaataataaggat 1716  
DB 1198 TAAATGATTTGATCTAAACATTAATTCATTAATATTTTCCAAATTAATATGTA 1257  
OY 1717 tcaagatcttgatgaaggtctgagcgagtgctataattccaccgaattcta 1776  
DB 1258 TAAATGATTTGATCTAAACATTAATTCATTAATATTTTCTAAATATGCT---- 1313  
OY 1777 acgagtttgaattttttttagcacaatcgactaactttgtacaaaatgaaagcta 1836

```

Db 1314 ----ATTATTGCATTAGTATATATTATTAATAGATTATTAATAAATATACCAT-ATTATCTA 1368
QY 1837 tatgtgaattttaaaaaaactcgacacataaagccgaagtaagtaaaatta 1896
Db 1369 TTTTACTAAAAAATACATAGACATTAATTAATACCATTTCTGCATTTTAAATTTT 1428
QY 1897 cctgcgaatttgcgaattcgccctcctaatttaatttttttggtgttaataatcg 1956
Db 1429 TATTGGAAATCTCTAATTATTATTTATTCATTATTATTATTAATTAATGTTTCTAGTTTAT 1488
QY 1957 gtatttactttaataaataaagtgagatgcatgataagcttggtgatalatag 2016
Db 1489 TTCAATACATTTTAAATATAATTTATTTATTTGGATATTATAGTATTAATATAT-TAAC 1547
QY 2017 agtgaatgtaatgacgatalattcttaaaaaaataaaaaa 2061
Db 1548 ATTGTGTTTAATATAGATTTTAAATTAATTAATAAATATA 1592

```

```

RESULT 12
US-07-991-867B-1
; Sequence 1, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruhl, Michael E.
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Sallwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/991,867B
; FILING DATE: 12-DEC-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Sallwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI14.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8457 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Amsacta moorei entomopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (65..1459)

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1474..2151
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (2239..2475)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2502..2987
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3080..6091
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (6277..6768)
; US-07-991-867B-1

```

Query Match 2.3%; Score 47; DB 1; Length 8457;

Best Local Similarity 46.2%; Pred. No. 0.014; Matches 270; Conservative 0; Mismatches 305; Indels 10; Gaps 3;

```

QY 1477 cttaatttgcgagttgttttcaaaatttaccatttgaatttgaattgagtaa 1536
Db 1018 CTCCTGTTTGGATATATCTGATTTAGAAAAGACTTCTCATATATATGATATGC 1077
QY 1537 attgactgaatttaactcgaaaataataaattcaattcgttgagacttcaat 1596
Db 1078 CTTTATATTTTATGATGTAATTAAGTATTAATATTATTATTCTGTAATATAT 1137
QY 1597 catcgattcccttaattgtctgtgcaaaattccatcatcgaattgaaaataatt 1656
Db 1138 TTTTATAGTCCAAATATGAAAAATTTCTTTAATATTAATTTTCAAAATTAAT 1197
QY 1657 caaatcgccaaatttgcgaacgaagaagataaattgcacaaataataaagat 1716
Db 1198 TATATGATTTGGATATATCTGATTTAGAAAAGACTTCTCATATATGATATGC 1257
QY 1717 tcaagtgatcctgataagatgctgcgagcgagttcatalattttccacggaattc 1776
Db 1258 TAAATGTTACTTTACCTCTGTTTCATCATCATCATATTTTCTTAATATAGCT---- 1313
QY 1777 acgagttttagaatttttttagccaaatcgacaaacttgcgaattgaaaatgta 1836
Db 1314 ----ATTATTGCATTAGTATATATTATTAATAGATTATTAATAAATATACAT-ATTATCTA 1368
QY 1837 tatgaataattttaaaaaaactcgacacataaagccgaagtaagtaaaatta 1896
Db 1369 TTTTACTAAAAAATTAACATAGACATTAATTAATACCATTTCTGCATTTTAAATTTT 1428
QY 1897 cctgcgaatttgcgaattcgccctcctaatttaatttttttggtgttaataatcg 1956
Db 1429 TATTGGAAATCTCTAATTATTATTTATTCATTATTATTATTAATTAATGTTTCTAGTTTAT 1488
QY 1957 gtatttactttaataaataaagtgagatgcatgataagcttggtgatalatag 2016
Db 1489 TTCAATACATTTTAAATATAATTTATTTATTTGGATATTATAGTATTAATATAT-TAAC 1547
QY 2017 agtgaatgtaatgacgatalattcttaaaaaaataaaaaa 2061
Db 1548 ATTGTGTTTAATATAGATTTTAAATTAATTAATAAATATA 1592

```

```

RESULT 13
US-08-544-332-1
; Sequence 1, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruhl, Michael E.
; TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77

```

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Gerard H. Bencen  
 STREET: 2421 N.W. 41st Street, Suite A-1  
 CITY: Gainesville  
 STATE: FL  
 COUNTRY: USA  
 ZIP: 32606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/544,332  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/991,867  
 FILING DATE: 07-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/107,755  
 FILING DATE: 19-AUG-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO 92/14818  
 FILING DATE: 12-FEB-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/827,685  
 FILING DATE: 30-JAN-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/657,584  
 FILING DATE: 19-FEB-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bencen, Gerard H.  
 REGISTRATION NUMBER: 35,746  
 REFERENCE/DOCKET NUMBER: UF114.C4  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 904-373-8100  
 TELEFAX: 904-372-5800  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 8457 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: Amsacta moorei entomopoxvirus  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: complement (65..1459)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1474..2151  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: complement (2239..2475)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 2502..2987  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 3080..6091  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: complement (6277..6768)  
 US-08-544-332-1

Query Match 2.3% Score 47 DB 4 Length 8457  
 Best Local Similarity 46.2% Pred. No. 0.014  
 Matches 270; Conservative 0; Mismatches 305; Indels 10; Gaps 3  
 QY 1477 cttaatttctgagctgttttcaaaaattatcttgtaattacattgagtgtaa 1536

Db 1018 CTCCTGTTTGGAFAATTCGATTGTTAGAAAAGACTTCCATATATAGGAATGC 1077  
 QY 1537 attagctgactttaaactgaaataataaattcaattcttgagactcttcaaat 1596  
 Db 1078 CTTTATATTTTATAGATGATATAAAGTAAATTTATATTTTCTGTAATAAT 1137  
 QY 1597 catctgattcccttaattctgctgtaaaattctcatcaatttgaaataaattc 1656  
 Db 1138 TTTTATAGTCCAAATAGAAAAATTTTCTTTAATATATTTTCAAAATATATAT 1197  
 QY 1657 caaatcgcacatttgaaccaaagaagataaatttgaccacaaataaagaagat 1716  
 Db 1198 TAATATGATTTGATCAATAAACATTAATTAATATATTTCCAAATATTTAGGTA 1257  
 QY 1717 tcaagtgacttgaatgaagctgcgagcgagttctatatttccaccgaattctta 1776  
 Db 1258 TAAATGTTACTTACCTCTGTTGATCATCATCATCTATTTTCTAATATACCT--- 1313  
 QY 1777 acgagtttgaatttctttagcacaatcgagctaacttgaacaaatgaagaagta 1836  
 Db 1314 ----ATATTGCAATTAATATATTTAATAGATTTAATAAATATACAT-ATTATCTA 1368  
 QY 1837 taigatgaattttaaacaacacacgaacaataaagaagccgaagtagtaaaatla 1896  
 Db 1369 TTTTACTAAAAAATTAACATGACATAAATTAATACAGATTGCGCATTTTAAATTTT 1428  
 QY 1897 ccgagcaaatctgcaattccgctcctatatttcaatttcttgctgcttaaaatcg 1956  
 Db 1429 TATTTGAAATCTTCTATTTATTTATTTATTTAATTAATTAAGTTCTGATTTAT 1488  
 QY 1957 gtaattacttaattaaataaagaatgaatgaatgaatgaatgaatgaatgaatga 2016  
 Db 1489 TTTCAATGATTTTAAATTAATTAATTTTATTTGATTTATAGATATTTATATAT 1547  
 QY 2017 agttgatgtaatgacgataatttcaaaaaaataaaaaaataaaaaa 2061  
 Db 1548 ATTGTGTTAATATATATATTTTAAATTAATTAATTAATTAATTAATTAATTA 1592  
 RESULT 14  
 US-08-913-374-1/C  
 Sequence 1, Application US/08913374  
 Patent No. 6057492  
 GENERAL INFORMATION:  
 APPLICANT: Petrus Theodorus Dehaan  
 TITLE OF INVENTION: Virus resistant or tolerant plants  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: No. 6057492artis Corporation  
 STREET: P.O. Box 12257  
 CITY: Research Triangle Park  
 STATE: No. 6057492Ch Carolina  
 COUNTRY: US  
 ZIP: 27709  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/913,374  
 FILING DATE: March 22, 1996  
 CLASSIFICATION: 800  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Hoxie, Thomas  
 REGISTRATION NUMBER: 32,993  
 REFERENCE/DOCKET NUMBER: 137-1099/PCT  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (919)541-8614  
 TELEFAX: (919)541-8689  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:



```

; LENGTH: 4821 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tosopovirus
; US-08-913-374-1

```

```

Query Match          2.2%; Score 45.6; DB 5; Length 4821;
Best Local Similarity 49.7%; Pred. No. 0.024;
Matches 145; Conservative 0; Mismatches 144; Indels 3; Gaps 1;

```

```

Qy 1764 caccgaattcctaagagtttgaatttttttagccaaatcgagacttaattgata 1823
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3800 CACTTATTAGCTTAATATTTCTGTATTTTTCATTTTGAATCCAAACCAACCAA 3741
Qy 1824 aaatgaagaattatgatgaattttaaaaaaacactcgacaataataagcccgaa 1883
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3740 AAACGAAACAAAAGAGAAAACAAAACAAAACAAAACAAAACAAAACAAAATAA 3681
Qy 1884 agtagtaaatctacgacgaatttgcaattcgccctcatttaatttttgg 1940
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3680 GCGCGAAAGCAAACTTTGCGCCGACAGACTTTTGTGTTTGTGTTTATTTTATTT 3621
Qy 1941 tctgttaataacgtctatcttactttaataaataaagtgagatgatagc 2000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3620 TTTTGTGTTTGTGTTTGTGTTTATTTATTTGTTTATTTATTTATTTATTTTAA 3561
Qy 2001 ttgttgatgatatagttgatgataatgataatcttcttaaaaaaa 2052
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3560 GATTTTCTATATATATATATCTGCTATATATAGAAATTTGAATTTTAA 3509

```

## RESULT 15

```

US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526tris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439

```

```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2

```

```

Query Match          2.2%; Score 45.4; DB 1; Length 5852;
Best Local Similarity 45.8%; Pred. No. 0.03;
Matches 232; Conservative 0; Mismatches 271; Indels 4; Gaps 2;

```

```

Qy 1479 ttaattgttcgagttgttttttcaaaattatctgttaatttacaattgagtgaaat 1538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2043 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTAA 1984
Qy 1539 ttgactgatttactgagaaataataatctatctgttgagactcttcaataatca 1598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1983 AGAAATAGAAAAGAGTTGGTAAACTACATAGTTTATATGTTTGTGATTTTAA 1924
Qy 1599 tctgatttccttactgtctgtgtcaaatctcatcaattggaataaataattcca 1658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1923 ATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1864
Qy 1659 aaat---cgltcaatttgaaccagaagaagataaaattgacccaataataaagga 1715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1863 AAATTTAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAG 1804
Qy 1716 ttcaatgatacttgatgaagtgctgtagcagagttctataattttccaccgaattct 1775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1803 TTCAATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1745
Qy 1776 aacgagttttgaatttttttagccaataatcgactaactttgacaaaatgaaaagt 1835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1744 AAAAATCTAGAAAATGAAATGTCATCAATAGTATTTTAAATTTTATTTTATTTT 1685
Qy 1836 atatgatgaattttaaaaaaacactcagacaataaagcccgaaagtagtaaat 1895
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1684 AAAAAGAAAAGTGTGATGCAAAAAAGTGTGATGCAAAAAAGTGTGATGCAAAAAAG 1625
Qy 1896 aactgacgaatttgcaatttcgcctcctatttaatttttgggtgttbaataatc 1955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1624 AGGGGAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1565
Qy 1956 ggtatattacttttaataataa 1892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1564 TTTAATTTGAATCATATGATTTACA 1538

```

Search completed: August 1, 2000, 18:17:13  
Job time: 74282 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:27:44 ; Search time 51.03 Seconds

(Without alignments)  
537.232 Million cell updates/sec

Title: US-09-147-955-4

Perfect score: 2337  
Sequence: 1 MYRRVLAATFPAQGHINPA.....TKLVRENATKTKTARQAMG 443

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1002	42.9	456	2	C71402
2	999.5	42.8	467	2	T02238
3	845	36.2	458	2	C71420
4	707	30.3	471	2	A54739
5	667	28.5	479	2	E71419
6	639.5	27.4	460	2	T00639
7	621.5	26.6	456	2	T00506
8	613.5	26.3	438	2	T00507
9	613.5	26.3	484	2	D71419
10	575.5	24.6	475	2	F71419
11	570.5	24.4	453	2	T00511
12	459.5	19.7	453	2	T45603
13	439	18.8	455	2	T00584
14	435	18.6	438	2	T45602
15	425	18.2	452	2	T12981
16	422	18.1	449	2	T45605
17	420.5	18.0	451	2	T45604
18	416.5	17.8	447	2	T12978
19	406.5	17.4	440	2	T00583
20	395.5	16.9	347	2	T06371
21	388.5	16.6	471	2	S01037
22	384	16.4	433	2	S51767
23	383.5	16.4	385	2	T01732
24	383	16.4	462	2	T01732
25	382.5	16.4	471	1	S08325
26	382.5	16.4	471	1	S01052
27	379.5	16.2	476	2	T03745
28	377	16.1	452	2	T00981
29	374.5	16.0	466	2	T07404

30	369.5	15.8	420	2	T08005	flavonol 3-O-glucosyl
31	368	15.7	488	2	T07786	UDP-glucose glucosyltransferase
32	364.5	15.6	476	2	T03747	glucosyltransferase
33	361	15.4	472	2	S39507	glucuronosyl transferase
34	356	15.2	452	2	G71416	probable glucosyltransferase
35	355	15.2	478	2	T05423	probable glucosyltransferase
36	354.5	15.0	478	2	T08395	UDP-glucose glucosyltransferase
37	350	15.0	487	2	S41951	UDP-glucose glucosyltransferase
38	349	14.9	455	1	X0BHF6	flavonol 3-O-glucosyltransferase
39	348	14.9	394	2	S41952	UDP-glucose glucosyltransferase
40	345	14.8	507	2	T46161	hypothetical protein
41	344.5	14.7	478	2	A71417	UDP-glucose glucosyltransferase
42	342.5	14.7	481	2	T01850	UDP-glucose glucosyltransferase
43	339.5	14.5	346	2	S41954	UDP-glucose glucosyltransferase
44	339	14.5	287	2	S41953	UDP-glucose glucosyltransferase
45	334.5	14.3	449	2	S41950	UDP-glucose glucosyltransferase

## ALIGNMENTS

RESULT 1	probable glucosyltransferase - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)	
A:Variety: Columbia	
C:Date: 03-Aug-1998	#sequence, revision 03-Aug-1998 #text, change 24-Nov-1999
C:Accession: C71402	
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D. P.; Medler, H.; Medler, E.; Wambolt, S.; Wetzenecker, T.; Pohl, T.M.; Terry, N.; Gavanagh, T.; Hempel, S.; Kotter, P.; Entlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B. Nature 391, 485-488, 1998	
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenach, A.; Jones, J.D.G.; Eneva, T.; Palmer, K.; Benes, V.; Rechman, S.; C.; Chalwalis, N.	
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana	
A:Reference number: A71400; MUID:98121113	
A:Accession: C71402	
A:Status: preliminary; nucleic acid sequence not shown; translation not shown	
A:Molecule type: DNA	
A:Residues: 1-456 <BEV>	
A:Cross-references: GB:597335; NID:g2244747; PID:e327440; PID:g2244766	
C:Genetics:	
A:Map position: 4COP9-4G3845	
C:Superfamily: Flavonol 3-O-glucosyltransferase	
Query Match	42.9% Score 1002; DB 2; Length 456;
Best Local Similarity	48.1% Pred. No. 3e-74; Mismatches 124; Indels 40; Gaps 13;
Matches 218; Conservative 71; Mismatches 124; Indels 40; Gaps 13;	
QY 3 RRRVLAATFPAQGHINPAALQFAKRLKAGTDVFTSYVARRMANFTASAAANPP--G 59	
DB 11 RPHYLTVFPAGHINPMLQLANRLIHGATVYTSTANSAARM-----GEPPSSK 62	
QY 60 LDPVAASGVDGDKPGDGKRYSEKAKSGSEALRNLLNND-----VTFVYSHL 112	
DB 63 LSFAMFTDGFDPDGKSPFDQIKYSELKRCGSMALNDIYKLANLATETETPTGYISVL 122	
QY 113 FAMAFAVARLTHVPALIMVPAATVLCYHRYEN-GYADEIDAGSNEIOLRLPSLEBS 171	
DB 123 VPMVSTARERHLPPTLMLIEPATVLDLYYFFNTSYKHLFDV--EPYKLRPLVITGGD 180	
QY 172 LPTFLLP--ATPERFLMKKEKLETLDEEKAVLVNTFDALPDALTAIDRYELIGCP 229	
DB 181 LPSFLQPSKALPSAL-VTLRHEIETALTESPNKILVNFSLLEHDAISVSKIMIPICP 239	
QY 230 LIPSAFLDGEDPSETSYGDLPEKSENNCYEMLSKSSSVYVYFSGS-VLRPPKAME 288	
DB 240 LVSSS--DGKT-----DLFKSSDE-DYTKWDSKLEKRSVYISLGTTHADDPKEMH 288	
QY 289 EIGKGLACGRPLIMIREQKNDGEEEEEELSCITELKMKKIVSWCSQLEVLAPRA 348	

Db 289 ALHGLVATNRPFELMIVREK---NPEKKKKRNFLELRGSDR-GLVVGWCOTAVLHCA 344  
 QY 349 LACFTVTHCGMSAVESLSCGIPVAVPQWFDQTNKLLIEDAMGCGVRVMEGGVDCG 408  
 Db 345 VCCFTTHCGMNSTLESLESGIPVAVPQWFDQTNKLLIEDAMGCGVRVMEGGVDCG 404  
 QY 409 EIERCEVYMDGDKTKLVRENAIKWKTILARQA 441  
 Db 405 EIRRCLEKVMGSGEEAEKRENAEKWAKAVDA 437

## RESULT 2

T02238  
 glucosyl transferase, jasmonate-induced - common tobacco  
 C:Species: Nicotiana tabacum (common tobacco)  
 C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 24-Nov-1999  
 C:Accession: T02238  
 R:Kojima, H.; Hashizume, K.; Imanishi, S.; Nakamura, K.  
 Submitted to the EMBL Data Library, January 1997  
 A:Description: Jasmonate-induced potential glucosyltransferase from tobacco suspension c  
 A:Reference number: 214633  
 A:Accession: T02238  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-467 <KOJ>  
 A:Cross-references: EMBL:AB00623; NID:d1094897; PID:d1019901  
 A:Experimental source: Strain BY-2  
 C:Genetics:  
 A:Gene: JIGT  
 C:Superfamily: flavonol O3-glucosyltransferase

Query Match 42.8%; Score 999.5; DB 2; Length 467;  
 Best Local Similarity 44.4%; Pred. No. 5e-74;  
 Matches 198; Conservative 83; Mismatches 144; Indels 21; Gaps 7;

QY 6 VLAFFPAOGHINPALQFAKRLKAGTDTFTSYVAMRMANNTASAAAGNPGLDFVAF 65  
 Db 10 VIALFFPGGHINPILQFASKLINLGKVTLSLSAFNRKIKLPK-----EGLTFAPF 64  
 QY 66 SGGYDGLKPGDGKR-YMSEKARSGEALRNLL-----LNDDVTFTVYSHLFAMAAYA 120  
 Db 65 SGGYDGNFSGFDYHLFNSAISKSHSEFIANLISKRAKNGYPPFFVITITLMDAGSVA 124  
 QY 121 RLSHVETALLMVEPATVLCIHYFENGYAD---EIDAGSENIOLPLPLDEGRSLPTFL 177  
 Db 125 KRLHIPSTLFWQOPATVFYIYRTFNFAFKNDSCQIIEPLPLSSDPFSFVAF 184  
 QY 178 PATPERFLM--MKERLETLDEGEKAKVLNTPDALPALTAIDRYELIGIPLIPSAF 235  
 Db 185 DVKSNDAVSEIKRQIEILINSEENRILVNTFDALNALRLKLVTVWVGIGLIPSSF 244  
 QY 236 LGEDESESYGGDLFEKSEENNCVEMLSKRSVYVYFSGVLAFFRAQMEIGKGL 295  
 Db 245 LDEKDKDNFFPADME--SENNYMEMLDARANKSVIYAFSVAEISSQWMEELISQGL 302  
 QY 296 ACGRFLMIRIADKNDGDEEEELSCIGELKKMGKIVSMCSOLELAHPALGCFVTH 355  
 Db 303 KCGRFLMIRIETLNG---EKPEEKLCKDELEKIGRIVRMCSEVYAKHSVGCFLTH 358  
 QY 356 CGMNSAVESLSCGIPVAVPQWFDQTNKLLIEDAMGCGVRVMEGGVDCGEIERCVA 415  
 Db 359 CGMNSTLESASGIPVAVPQWFDQTNKLLIEDAMGCGVRVMEGGVDCGEIERCVA 418  
 QY 416 NYMDGDKTKLVRENAIKWKTILARQA 441  
 Db 419 IVMGDAEEGELRKNAQWKDLAKES 444

## RESULT 3

C71420  
 hypothetical protein - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)

A:Variety: Columbia  
 C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999  
 C:Accession: C71420  
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D  
 P.; Wedler, H.; Medler, E.; Wandut, R.; Welzenegger, T.; Pohl, T.M.; Terry, N.; G  
 vanagh, T.; Hempel, S.; Kotter, P.; Ertlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B  
 Nature 391, 485-488, 1998  
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen  
 erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;  
 C:Chalvatzis, N.  
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t  
 A:Reference number: A71400; MUID:98121113  
 A:Accession: C71420  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-458 <BEV>  
 A:Cross-references: GB:Z97339; NID:g2244901; PID:e326935; PID:g2244911  
 C:Genetics:  
 A:Map position: ACOP9-4G3845  
 C:Superfamily: flavonol O3-glucosyltransferase

Query Match 36.2%; Score 845; DB 2; Length 458;  
 Best Local Similarity 41.7%; Pred. No. 2.2e-61;  
 Matches 194; Conservative 74; Mismatches 129; Indels 68; Gaps 17;

QY 7 LLAFFPAOGHINPALQFAKRLK--AGTDTFTSYVAM--RRANNTASAAAGNPGLDFV 63  
 Db 15 LFTVTPAOGHINPILQFASKLINLGKVTLSLSAFNRKIKLPK-----EGLTFAPF 70  
 QY 64 AFSDDYDGLKPG--GDGR-----YMSEKARSGEALRNLL-----LNDDVTFTVYSH 111  
 Db 71 TYSDDGDDGFKSAYSDKRODATGNFMSEMRKKEITLIEDRKQNRPTCVVYIT 130  
 QY 112 LRMAAEVRLSVLPALLMVEPATVLCIHYFENGYADEI---DAGSENIOLPLPL 167  
 Db 131 LTLVAEELAFS-----IFYHFNCEYDAISEMANTPSSIKPLPL 174  
 QY 168 EGRSLPTFLPATPERFLR--MKERLETLDEGEKAKVLNTPDALPALTAIDRYELI 225  
 Db 175 TVRDLPSTVSSVAVAFLLPATREQIDSLKEELNPKILNTPDELEPMSSVPDFKIV 234  
 QY 226 GIGPLIPSAFLGDEPSETSYGGDLFEKSEENNCVEMLSKRSVYVYFSGVLAFFRA 285  
 Db 235 PVGPPL-----TLRDF-----SRGEIEMLDTRKADSVLYVSGFLAVLSK 278  
 QY 286 QMEIKGKLACGRPLMIRIADKNDGDEEEELSCIG---ELKMGKIVSMCSQ 341  
 Db 279 OLVELCKALIQSRPPLMWTTRKSYRNKEDQEKE--DCISFREBLDEIGMVSQDF 337  
 QY 342 EYLAHPALGCFVTHCGMNSAVESLSCGIPVAVPQWFDQTNKLLIEDAMGCGVRV--RM 399  
 Db 338 RYLNHNSICGFTVTHCGMNSTLESLSGVPVAVPQWFDQTNKLLIEDAMGCGVRVMEK 397  
 QY 400 NREG--GVDGEIERCEVYMDGDKTKLVRENAIKWKTILARQA 442  
 Db 398 EERGVAVVDSEIRIRCEIYME--DKAEFRGNATRMKDLAAEAV 440

## RESULT 4

A54739  
 indole-3-acetate beta-glucosyltransferase (EC 2.4.1.121) - maize  
 C:Species: Zea mays (maize)  
 C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 29-Oct-1999  
 C:Accession: A54739  
 R:Szerszen, J.B.; Szczygiowski, K.; Bandurski, R.S.  
 Science 265, 1699-1701, 1994  
 A:Title: Iaglu, a gene from Zea mays involved in conjugation of growth hormone indole  
 A:Reference number: A54739; MUID:94367368  
 A:Accession: A54739  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-471 <SZS>



```

221 RYELIGIPLPASFLDGEDPSETSYGDLFEKSENNCEVEMINSKPKSSVYVSGSVL 280
232 -----TLIGPPIPAVLDDEDDKDGASLL-KPISKECHEMETKQASVAFVSGSFG 286
QY 281 RPKAKMEIEIGKGLACGPFPLMIRE--QKNDGEEEBEELSCIGELKKMKITVSW 337
Db 287 ILFEKOLAEVAIALQESDLPLFWIKFAHIAKLPEGFVE-----STKDRALLVSW 336
QY 338 CSOLEVLAHAFALCGFVTHCGMNSAVESLSCIGIVVANPQMFDDOTNAKLIEDAMGCVRV 397
Db 337 CNOLEVAHESICGFLTHCGMNSTLEGLSGVPMVGPQMSDDMNDAKFVEEYWKGYRA 396
QY 398 RMEGGG-VDCIEIERCEVEMVMDGDKTKLVRENAIKMTLAROAM 442
Db 397 KEBAGEIVASEELVRLCKVMEGESSVK-IRRESSKMKMLAVAKAM 441

RESULT 7
T00506
Indole-3-acetate beta-glucosyltransferase homolog T20D16.11 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00506
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, November 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.
A:Reference number: 214159
A:Accession: T00506
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-456 <R0U>
A:Cross-references: EMBL:AC002391; NID:92642427; PID:92642451
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Note: T20D16.11
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 26.6%; Score 621.5; DB 2; Length 456;
Best Local Similarity 34.7%; Pred. No. 4.2e-43;
Matches 156; Conservative 86; Mismatches 171; Indels 37; Gaps 14;

QY 6 VLATPPAGCHINPALQFAKRLKAGTDVTF-FTSVYAMRRMANTASAAAGNPGLDFVA 64
Db 11 VLWATLPFGCHINPLAKHLKHLSSKNLHINLATIESADLSTVAKPR---YVPLVLF 67
QY 65 FSDGYDDGLKPGDGKRYNEMKARGSEALRNL-LINNDVTFVYVSHLFAMAAYEARL 122
Db 68 FSDGL-----PKEDPKAPFLLKSLNKVGAANLSKIIEEKRYSCTIISSPPTPVWPAVAAS 122
QY 123 SHVPTALMWEPATVLCIYHFF--NGYADEIDAGSNEIOLPRLPSLEORSLEPTFLPA 179
Db 123 HNSICALMQLQAGAVSYRYRYMKNTPDLDEL-NQYELPALPLEVNDLPSEFMLS 181
QY 180 TPRFRRLMKEKLETLDEGEKAKLVNTPDALBPDLTAI-DREYELGIGPLIPSAFLDG 238
Db 182 GGAHFYLMALBPADCL-RYVKVLYVNSFYELSEIIESMADLKVPYIPGLV-SPFLIG 238
QY 239 EDPSETSYGDLFEKSENNCEVEMINSKPKSSVYVSGSVLFRPKAKMEIEIGKGLACG 298
Db 239 DGEETFLDGKNDLFCSSDDCCMEMLDKQARSVYVIFSGMLLETLENQVETIAKALKNRG 298
QY 299 RPLLMIREKNDGDEEBEELSCIGELKKMK--IYWCSOLEVLAHPALGCFVTHC 356
Db 299 LPLLMYIR-----PKERAKQNVAVLOEYVKEGGVLEMSPOEKLITSHAIISCFVTHC 350
QY 357 GMSAVESLSCIGIPVAVPQMFDDOTNAKLIEDAMGCVVRNMEGGGVG-----CEIER 412
Db 351 GMSVETVETVAGVYVAVPMTQDPIDARLLVDFVIGVGRM--NDSVDELKVEEVER 407
QY 413 CVMVMGDKDKTLVRENAIKMTLAROAM 442
```

```

Db 408 CIEAVTEGPAAYD-IRRAAEKLRVARLAL 436

RESULT 8
T00507
Indole-3-acetate beta-glucosyltransferase homolog T20D16.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00507
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, November 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.
A:Reference number: 214159
A:Accession: T00507
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-438 <R0U>
A:Cross-references: EMBL:AC002391; NID:92642427; PID:92642438
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Note: T20D16.12
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 26.3%; Score 613.5; DB 2; Length 438;
Best Local Similarity 34.8%; Pred. No. 1.8e-42;
Matches 154; Conservative 86; Mismatches 158; Indels 45; Gaps 18;

QY 14 QGHINPALQFAKRLKAGTDVTFFTSVYAMRRMANTASAAAGNPGLDFVAFSDGY--DD 71
Db 7 QGHINPLKRAKHLARNLHFTLATTEQADLSTADEP--HRP-VDLAFSDGLPKDD 63
QY 72 GLKPGDGKRYNEMKARGSEALRNLLNN--DDVTFVYVSHLFAMAAYEARLSHVPTAL 129
Db 64 PRDPTLAK-----SLKDKGAKNLSKIIEEKRFDDCIISVPT--PWPVAAAANINICAI 116
QY 130 LWEVPATVLCIYHFF--NGYADEIDAGSNEIOLPRLPSLEORSLEPTFLPAPRFRRL 186
Db 117 LMTQACAFSVYRYRYMKNTPDLDEL-NQYELPALPLEVNDLPSEFMLS 175
QY 187 MKKEKLETLDEGEKAKLVNTPDALBPDLTAI-DREYELGIGPLIPSAFLDGEDPSETS 245
Db 176 IMAEFADCL-KDVKVLYVNSFYELSEIIESMSDLKPIIPGLV-SPFLGNDDEKTL 232
QY 246 YGGDLFEKSENNCEVEMINSKPKSSVYVSGSVLFRPKAKMEIEIGKGLACGPFPLMI 305
Db 233 ---DMWK--VDDYCMEMLDKQARSVYVIFSGITLSENOVETIAVALNRGVPLMVI 287
QY 306 REQKNDGEEEBEELSCIGELKKMK--IYWCSOLEVLAHPALGCFVTHCGMNSAVE 363
Db 288 R-----PREKENQVLOEYVKEGGVLEMSPOEKLITSHAIISCFVTHCGMNSTIE 339
QY 364 SLSCIGIPVAVPQMFDDOTNAKLIEDAMGCVVRNMEGGGVG-----CEIERCEVEM 419
Db 340 TVYTGVPVAVPMTQDPIDARLLVDFVIGVGRM--NDAISBELVAVERCIEAVTE 396
QY 420 GGDGKTLVRENAIKMTLAROAM 442
Db 397 GPAADW-RRRATLKHAAARSAM 418

RESULT 9
D71419
Probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
C:Accession: D71419
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
P.; Medler, H.; Medler, E.; Wamboldt, R.; Weltenssger, T.; Pohl, T.M.; Terryn, N.; G
avanagh, T.; Hempel, S.; Kotter, P.; Entlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
```

Nature 391, 485-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech et al., A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans

C.: Chalwatzis, N.

A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal

A:Reference number: A71400; MUID:98121113

A:Accession: D71419

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-484 <BEV>

C:Genetics:

A:Cross-references: GB:J297339; NID:g2244901; PID:e327472; PID:g2244905

A:Map position: 4COP9-4G3845

C:Superfamily: Flavanol O3-glucosyltransferase

Query Match 26 3%; Score 613.5; DB 2; Length 484;

Best Local Similarity 33.2%; Pred. NO.2.1e-42;

Matches 154; Conservative 90; Mismatches 171; Indels 49; Gaps 15;

QY 6 VLLATPEAGSHINPALQFAKRLKAGTVDVFETTSYAAVRMANNTASAAAG--NPPG---L 60

DB 14 VMLVSPGGQGHVAPLRLKGLLAKSGLLVTFVTELMCKMKMQAKIYDGLKPVSSGI 73

QY 61 DVAESDGDYDGLKPGGDKRYMSEMKARG---SEALRNLLNDDVTFVYSHLFAMA 116

DB 74 REFFPEEWEAEDDDRRADFSLYIAHLESVGIRESKLVRYEANEPEVCLINPFIPWY 133

QY 117 AEVARLSHVPTALLWVEPAVLCYHFFENGAYD-----EIDAGSNIOQLPRLSLEQ 169

DB 134 CHVAEERNIPICAVLWVOSCAFSAFYHYDGSVSEFTEPELDD---VKLPVVLKN 188

QY 170 RSLPTFLPATP-ERFELMKKEKLETLDEEKAKVLTNTFALPDAITADRY-ELIGI 227

DB 189 DEPSFLHPSRRTGFGQALIGQFKNL--SKSPCLIDISFDLDEVIDYMSSLCVKTIV 246

QY 228 GPLIPSAFLDGEDPSETSYGGDLFEKSENNOCVEMLNKPKSSVYVSFGSVLRFKRAQM 287

DB 247 GPLFKARFVTSVVS-----GDICKSTDK--CLEMLDSRPKSSVYISFGTAYLKEOQI 299

QY 288 EELGKGLIACGRPLMKIRQGNKDDGE-----EEREELSCIGELKMKKRTIVSMCSOLEY 343

DB 300 EELAHGVLSGSLSFVLMYIRPPHDLKVTETHVLPOLKLTSSASAGK---GMIVDMCPQEOY 355

QY 344 LAHPALGCEFTHCAGMNSAVESLSCGIPVAVAPQWEDQTNNAKLIEDAMGTGVRVNNEG 403

DB 356 LSHRVSACFTTHCGMNSTMESLSSGVPVCCQPMQDQYTDAYLLIDVFETGYAL-----GR 411

QY 404 GVDGCEI---ERCVEWYMDG--GDKTKLVRENAIKWTKLARQAM 442

DB 412 GATEERNVPRREEVAERKLEATVGEKAELRKRLKALMKAEAAV 455

RESULT 10

F71419

Probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

A:Variety: columbia

C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999

C:Accession: F71419

R:By: M.; Hancock, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk

P.; Medler, H.; Medler, E.; Wambutt, R.; Meltzner, T.; Poll, T.M.; Terry, N.; Glei

avangh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.

Nature 391, 485-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech et al., A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans

C.: Chalwatzis, N.

A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal

A:Reference number: A71400; MUID:98121113

A:Accession: F71419

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-475 <BEV>

A:Cross-references: GB:297339; NID:g2244901; PID:e326932; PID:g2244907  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: Flavonol O3-glucosyltransferase

Query Match      24.6%, Score 575.5; DB 2; Length 475;  
Best Local Similarity 31.8%; Pred. No. 2,6e-39;  
Matches 148; Conservative 83; Mismatches 176; Indels 59; Gaps 16;

OY    6 VLATFPAGGHINPALQRAKLKAGTDVTFYSYAMRRANNTASAAANGPGLDFA 62  
     |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
DB    10 VMLVFSPGGCHISPLRIGKIITASKGIIVTFTVTTEEPRLGKKRKARANNIQDGVLPGVGIF 69  
     :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
OY    63 VAFSGDYGDLKPFGDDGRKNSEMKARGSEALRNLL--LNNDDVTFPVYSHLFAMAEEVA 120  
     LRF-EFFEGGFYYKKEDFDLLQKSLEVSCKRETKNLVKTYEKQPRLCNNAFVRWCDIA 128  
     ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
DB    70 LRF-EFFEGGFYYKKEDFDLLQKSLEVSCKRETKNLVKTYEKQPRLCNNAFVRWCDIA 128  
     :||:||||:|::|::|::|::|::|::|::|::|::|::|::|::|::  
OY    121 RLSHVPALLNVEPAATVCITYHFYPNGYADEIADGSNEI--QLPRLP-SLEORSLPYFL 177  
     ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
DB    129 EELQIPSAVALWQSACLAAYYYYHHOLVKPTPEPETEPIVDVPFKPLTLKHDEIPSFLH 188  
     :||:||||:|::|::|::|::|::|::|::|::|::|::|::|::|::  
OY    178 PATPERFRLMKEKETLEDGE--EKAK-----VLVTPALRPDALTAID----RYEL 224  
     |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
DB    189 PSSP-----LSSIGGTLEGIKRLMPSYLFIETFECLKRDVIDHSQCLPNVNF 238  
     :||:||||:|::|::|::|::|::|::|::|::|::|::|::|::|::  
OY    225 IGILPLPSAFLDGEDPSTSYSGDLPFEEKSEENNCVELANKPKRSSVVYVSFGSVLRPPK 284  
     ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
DB    239 NPICGLEPTMA-----KTIRSDIKGDI--SKPDSCIEHLDRBPRESSVYIIISFGIALFLKQ 291  
     ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
OY    285 AQMEIIGNKLGACGPRTLMIREQRNDKGDEEEEEELSCTIGELKKMGKYSMCSOLEVL 344  
     ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
DB    292 NCIDEIARIHGINSGLTWLTAPPL---EGLAIPEHYLP-L-ELEEKGKIVEMCOQEKEYL 346  
     :||:||||:~::~|::|::|::|::|::|::|::|::|::|::|::|::|::  
OY    345 AHPALCGCFTHGCMSNAVESLCGIPPVAVAQMPQTNALIEDANTGYRVVMNEGCG 404  
     ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
DB    347 AHPAVCAECFHSCGMNSTMEDALTGVVICPCPWGQQVTVANAAYMDIVFYTGRLSR---- 401  
     ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
OY    405 VGCEIERCV-----EWAMDG--GDKTKLVRENAILKKTKLARQA 442  
     |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
DB    402 --GASDERLYPREVEAEERILEATTVEKADELRENARKKEEFESAIV 445  
     |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::

RESULT 11  
T00S11  
Indole-3-acetate beta-glucosyltransferase homolog T20D16.16 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999  
C:Accession: T00S11  
R:Roundley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K submitted to the EMBL Data Library, November 1997  
A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.  
X:Reference number: Z14159  
A:Accession: T00S11  
A>Status: translated from GR/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-453 <ROU>  
A:Cross-references: EMBL:AC002391; NID:g2642427; PIDs:g2642442  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 2  
A:introns: 266/2  
A>Note: T20D16.16  
C:Superfamily: flavanone O3-glucosyltransferase

Query Match      24.4% ; Score 570.5; DB 2; Length 453;  
Best local similarity 33.3%; Pred. No. 6,2e-39;  
Matches 153; conservative 77; mismatches 162; indels 67; gaps 17;

OY    6 VLATFPAGGHINPALQRAKLKAGTDVTFYSYAMRRANNTASAAANGPGLDFA 64  
     |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
DB    11 VMALVAPFGCHINPMILKFAPHKHAIRTNLHTLATISSADDLSLT-----DFPHSIYDVLF 65







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:31:33 ; Search time 61.65 seconds  
(without alignments)  
498.216 Million cell updates/sec

Title: US-09-147-955-4

Perfect score: 2337

Sequence: 1 MVRRLVLLATPAQGHINPA.....TKLVRENAIKWKTLLAQAMG 443

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_12:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2337	100.0	443	10 Q9ZR26	Q9ZR26 perilla fru
2	2223	95.1	460	10 Q9ZR27	Q9ZR27 perilla fru
3	1583.5	67.8	461	10 Q9ZR25	Q9ZR25 verbena x h
4	1002	42.9	456	10 Q9Z270	Q9Z270 arabidopsis
5	999.5	42.8	467	10 P93709	P93709 nicotiana t
6	940	40.2	519	10 Q9ZVY2	Q9ZVY2 arabidopsis
7	923	39.5	474	10 Q04930	Q04930 arabidopsis
8	914.5	39.1	455	10 Q9ZVY5	Q9ZVY5 arabidopsis
9	845	36.2	458	10 Q23406	Q23406 arabidopsis
10	667	28.5	479	10 Q23401	Q23401 arabidopsis
11	657.5	28.1	449	10 Q22822	Q22822 arabidopsis
12	644.5	27.6	449	10 Q22820	Q22820 arabidopsis
13	639.5	27.4	460	10 Q48676	Q48676 arabidopsis
14	621.5	26.6	456	10 Q22182	Q22182 arabidopsis
15	613.5	26.3	436	10 Q22183	Q22183 arabidopsis
16	613.5	26.3	484	10 Q23400	Q23400 arabidopsis
17	575.5	24.6	475	10 Q23402	Q23402 arabidopsis
18	570.5	24.4	453	10 Q22186	Q22186 arabidopsis
19	551	23.6	481	10 Q9ZVJ3	Q9ZVJ3 arabidopsis

20	468	20.0	482	10 Q9ZUV0	Q9ZUV0 arabidopsis
21	439	18.8	455	10 Q64733	Q64733 arabidopsis
22	426.5	18.2	477	10 Q9ZWQ5	Q9ZWQ5 vigna mungo
23	406.5	17.4	440	10 Q64732	Q64732 arabidopsis
24	399.5	17.1	455	10 Q9ZWS2	Q9ZWS2 vigna mungo
25	396.5	17.0	491	10 Q9ZQ99	Q9ZQ99 arabidopsis
26	395.5	16.9	347	10 Q49042	Q49042 pisum sativ
27	395.5	16.9	496	10 Q9ZQ97	Q9ZQ97 arabidopsis
28	388	16.6	496	10 Q9ZQ96	Q9ZQ96 arabidopsis
29	383	16.4	454	10 Q9XF16	Q9XF16 forsythia x
30	383	16.4	462	10 Q04622	Q04622 arabidopsis
31	383	16.4	495	10 Q9ZQ95	Q9ZQ95 arabidopsis
32	382	16.3	496	10 Q9ZQ98	Q9ZQ98 arabidopsis
33	380	16.3	484	10 Q9ZQ94	Q9ZQ94 arabidopsis
34	379.5	16.2	476	10 P93364	P93364 nicotiana t
35	377	16.1	452	10 Q48715	Q48715 arabidopsis
36	374.5	16.0	466	10 Q43526	Q43526 lycopersico
37	374.5	16.0	474	10 Q82382	Q82382 arabidopsis
38	374	16.0	447	10 Q04114	Q04114 perilla fru
39	373.5	16.0	495	10 Q9ZQ94	Q9ZQ94 arabidopsis
40	369.5	15.8	420	10 Q22603	Q22603 ipomoea pur
41	368.5	15.8	280	10 Q9ZMQ4	Q9ZMQ4 vigna mungo
42	368	15.7	488	10 P93789	P93789 solanum tub
43	367.5	15.7	481	10 Q82381	Q82381 arabidopsis
44	364.5	15.6	476	10 P93365	P93365 nicotiana t
45	359	15.4	381	10 Q9ZMQ3	Q9ZMQ3 vigna mungo

#### ALIGNMENTS

RESULT 1

Q9ZR26 ID Q9ZR26 PRELIMINARY; PRT; 443 AA.

AC Q9ZR26;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE HOMOLOGUE.  
GN PF3R6.

OS Perilla frutescens.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
core eudicots; Asteridae; euasterids I; Lamiales; Lamiales; Lamiaceae; Perilla.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. SHIKUN; TISSUE-LEAF;

RX MEDLINE; 99167509.

RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,

RA KUSUMI T., SAITO K.;

RT "Molecular cloning and biochemical characterization of a novel  
anthocyanin 5-O-glucosyltransferase by mRNA differential display for  
plant forms regarding anthocyanin.";  
RL J. Biol. Chem. 274:7405-7411(1999).

DR EMBL; AB013597; BAA36422.1; -

DR PROSITE; PS00375; UDPGT; 1.

KW Transférase.  
SQ SEQUENCE 443 AA; 49110 MW; D87B5486 CRC32;

Query Match 100.0%; Score 2337; DB 10; Length 443;  
Best Local Similarity 100.0%; Pred. No. 2.1e-185;  
Matches 443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVRRLVLLATPAQGHINPALQFAKRLKAGTDVFTFTSVYAWRRMANTASAAAGNPGL 60

Db 1 MVRRLVLLATPAQGHINPALQFAKRLKAGTDVFTFTSVYAWRRMANTASAAAGNPGL 60

Qy 61 DFVAFSDGVDGLKPGDGKRYMSEMKARGSEALNLLNDDVTFVYVSHLFAWAEVA 120

Db 61 DFVAFSDGVDGLKPGDGKRYMSEMKARGSEALNLLNDDVTFVYVSHLFAWAEVA 120

Qy 121 RLSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNEIQLPRLPSLEQRSPLFLPAT 180

```

Db 121 RLSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNEIQLPRLPSLEQSLPTFLPAT 180
QY 181 PERFLRMMEKLETLGEEKAKVLVNTFDALPDALTAIDRYELIGIGPLIPSAFLDGED 240
Db 181 PERFLRMMEKLETLGEEKAKVLVNTFDALPDALTAIDRYELIGIGPLIPSAFLDGED 240
QY 241 PSTSYGGDLFEKSEENNCVWLNKSPKSSVYVVSFGSVLRFPPKAQMEETGKLLACGRP 300
Db 241 PSTSYGGDLFEKSEENNCVWLNKSPKSSVYVVSFGSVLRFPPKAQMEETGKLLACGRP 300
QY 301 FLWMIREQKNDGDEEEEBELSCIGELKMKGIKIVSWCSQLEVLHAPALGCFVTHCGWNS 360
Db 301 FLWMIREQKNDGDEEEEBELSCIGELKMKGIKIVSWCSQLEVLHAPALGCFVTHCGWNS 360
QY 361 AVESLSGIPVAVPOWFOQTNAKLIEDAWGTGVRVRNNEGGVDCETERCERCEVAVMDG 420
Db 361 AVESLSGIPVAVPOWFOQTNAKLIEDAWGTGVRVRNNEGGVDCETERCERCEVAVMDG 420
QY 421 GDKTKLVRENAIKWKTARQAMG 443
Db 421 GDKTKLVRENAIKWKTARQAMG 443

RESULT 2
Q92R27 PRELIMINARY; PRT; 460 AA.
AC Q92R27;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE.
GN PF3R4.
OS Perilla frutescens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Lamiales; Perilla.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, SHIKUN; TISSUE-LEAF;
RX MEDLINE; 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL; AB013596; BAA36421.1; -.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 460 AA; 50974 MW; FE7CCF22 CRC32;

```

```

Query Match 95.1%; Score 2233; DB 10; Length 460;
Best Local Similarity 95.7%; Pred. No. 5.9e-176;
Matches 424; Conservative 9; Mismatches 8; Indels 2; Gaps 1;

QY 1 MVRRLVLLATFPAQGHINPALQFAKRLKLLKAGTDVTFSTSYAARRKANTASAAAGNPPGL 60
Db 1 MVRRLVLLATFPAQGHINPALQFAKRLKLLKAGTDVTFSTSYAARRKANTASAAAGNPPGL 60
QY 61 DFVAFSDGYDGLKPGDGKRYMSEMKARGSEALRNLLNNDVTFVYSHLFAWAAEVA 120
Db 61 DFVAFSDGYDGLKPGDGKRYMSEMKARGSEALRNLLNNDVTFVYSHLFAWAAEVA 120
QY 121 RLSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNEIQLPRLPSLEQSLPTFLPAT 180
Db 121 RESQVPSALLWVEPATVLCIYHFYFNGYADEIDAGSNEIQLPRLPSLEQSLPTFLPAT 180
QY 181 PERFLRMMEKLETLGEEKAKVLVNTFDALPDALTAIDRYELIGIGPLIPSAFLDGED 240
Db 181 PERFLRMMEKLETLGEEKAKVLVNTFDALPDALTAIDRYELIGIGPLIPSAFLDGED 240

```

```

QY 241 PSTSYGGDLFEKSEENNCVWLNKSPKSSVYVVSFGSVLRFPPKAQMEETGKLLACGRP 300
Db 241 PSTSYGGDLFEKSEENNCVWLNKSPKSSVYVVSFGSVLRFPPKAQMEETGKLLACGRP 300
QY 301 FLWMIREQKNDGDEEEEBELSCIGELKMKGIKIVSWCSQLEVLHAPALGCFVTHCGWNS 360
Db 301 FLWMIREQKNDGDEEEEBELSCIGELKMKGIKIVSWCSQLEVLHAPALGCFVTHCGWNS 360
QY 361 AVESLSGIPVAVPOWFOQTNAKLIEDAWGTGVRVRNNEGGVDCETERCERCEVAVMDG 420
Db 361 AVESLSGIPVAVPOWFOQTNAKLIEDAWGTGVRVRNNEGGVDCETERCERCEVAVMDG 420
QY 421 GDKTKLVRENAIKWKTARQAMG 443
Db 421 GDKTKLVRENAIKWKTARQAMG 443

RESULT 3
Q92R25 PRELIMINARY; PRT; 461 AA.
AC Q92R25;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE.
GN HGT8.
OS Verbena x hybrida.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Verbenaceae;
OC Verbena.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PETAL;
RX MEDLINE; 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL; AB013598; BAA36423.1; -.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 461 AA; 51347 MW; 6DCBC3B6 CRC32;

Query Match 67.8%; Score 1583.5; DB 10; Length 461;
Best Local Similarity 68.7%; Pred. No. 4.8e-123;
Matches 310; Conservative 57; Mismatches 65; Indels 19; Gaps 7;

QY 1 MVRRLVLLATFPAQGHINPALQFAKRLKLLKAGTDVTFSTSYAARRKANTASAAAGNPPGL 60
Db 1 MVRRLVLLATFPAQGHINPALQFAKRLKLLKAGTDVTFSTSYAARRKANTASAAAGNPPGL 60
QY 61 DFVAFSDGYDGLKPGDGKRYMSEMKARGSEALRNLLNNDVTFVYSHLFA 114
Db 61 DFVAFSDGYDGLKPGDGKRYMSEMKARGSEALRNLLNNDVTFVYSHLFA 114
QY 115 WAAEVARLSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNEIQLPRLPSLEQSLP 173
Db 115 WAAEVARLSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNEIQLPRLPSLEQSLP 173
QY 174 TFLPATPERFLRMMEKLETLGEEKAKVLVNTFDALPDALTAIDRYELIGIGPLIPS 233
Db 174 TFLPATPERFLRMMEKLETLGEEKAKVLVNTFDALPDALTAIDRYELIGIGPLIPS 233
QY 234 AFLDGEDPSTSGDGLFEK-SEENNCVWLNKSPKSSVYVVSFGSVLRFPPKAQMEETG 292
Db 234 AFLDGEDPSTSGDGLFEK-SEENNCVWLNKSPKSSVYVVSFGSVLRFPPKAQMEETG 292
QY 293 GLLACGRPFLLMIREQKNDGDEEEEBELSCIGELKMKGIKIVSWCSQLEVLHAPALGCF 352
Db 293 GLLACGRPFLLMIREQKNDGDEEEEBELSCIGELKMKGIKIVSWCSQLEVLHAPALGCF 352

```

Query Match	42.8%	Score	999.5	DB	10	Length	467
Best Local Similarity	44.4%	Pred. No.	1e-74				
Matches	198	Conservative	83	Mismatches	144	Indels	21
Gaps	7						
Qy	6	VLLATFPAQGHINPALQFAKELLKAGTDVTFFTSVYARRMANTASAAAGNPPGLDFVAF	65				
Db	10	VLIAFFQGHINPSLOFSKLLNLGVKVTLUSSLSAFNRKLNPKI-----EGLTFAPF	64				
Qy	66	SDGVDDGLKPGGDKR-YMSEMKARGSEALRNLL-----LNNDVTFVYVYSHLFAWAAREVA	120				
Db	65	SDGVDDGNFKGSFDDYHLFNLSAIIKSHGSEFIANLIKSAKNAGYPTFRVIYITLMDWAGSVA	124				
Qy	121	RLSHVPTALLMVEPATVLCIYHFYNGYAD---EIDAGSNEIQIPRLPSLSQRSUPLTELL	177				
Db	125	KKLHPSTLEWIQPATVFDIYYIRFTNFANYFKNYDSODQIIEPLGPLSSSDFFPSVF	184				
Qy	178	PATPERFLM--MKELLETLDGEKAKVLVNTFDALPDALTADRVELIGIGLPLPSAF	235				
Db	185	DDVKSNDWAVESIAKQIEILNSEENPRILVNTFDALNLALRVUKNTVWGIGLPLPSF	244				
Qy	236	LDGEDPSETSYYGGDLFEKSEENNVCEMLNPKSPSVVYVSGFSLRPFKAQMEIRGKLL	295				
Db	245	LDERDKDNFFAADWIE--SENNTMWEMLDRANKSVIYAFGSYAEISSQWMEISQGLL	302				
Qy	296	ACGRPFTLWIREQKNDGDEBEEREEELSCIGELKMGKIVSWCSQLEVLAIHPALGCPVTH	355				
Db	303	KCGRPFTLWIRETLNG----EKPEKUTCKDELKIGRIVWCQMEYVKESSGYEGLTH	358				



```

DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE T25N20.17.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,
RA VISOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RT *Genomic sequence for Arabidopsis thaliana BAC T25N20.*;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005106; AAC80596.1; -.
DR PROSITE; PS00375; UDPGT; 1.
SQ SEQUENCE 455 AA; 51190 MW; F9BAC33E CRC32;

Query Match 39.1%; Score 914.5; DB 10; Length 455;
Best Local Similarity 43.2%; Pred. No. 1e-67;
Matches 196; Conservative 86; Mismatches 141; Indels 31; Gaps 11;

QY 1 MVRRLVLLATPPAAGHINPALQFAKRLK--AGTDVTFSTSYAWRRMANTASAAAGNPPG 59
Db 1 MAQPHFLVTFPAAGHVNPSLRFAARLIKTTGARVTFATCLSVIHR---SMIPNNHNVEN 57
QY 60 LDFVAFSDGDDG--LKPGGDKRYMSEMKARGSEALRNLLINDD----VTFVYVSHLFA 114
Db 58 LSLFTFSDGDDGVISNTDDVQNLVHFERNQDKALSDFEANQNGDSPVSCLIITLPLN 117
QY 115 WAAEVARLSHVPTALLWVEPATVLCIYHFYNGVADIDAGSNEI--QLPRLPSLEQRS 173
Db 118 WVPKVARFRPLSVHLWQAFADFIYNY-----STGNNSVFEPNLPISLEIRDL 169
QY 174 TFLPLA--TPERFLMKKEKLETGEEKAKVLVNTFDALPDALTAIDRYELIGILP 232
Db 170 SFLSPNTNAAQVYQELMDFLKEENPKILVNTFDLSPEFTALPNTNEMAVGFLP 229
QY 233 SAFLDGEDPSETS YGGDLFEKSENNVEMLSNPKSVVYSGVSRFPKQMEBIG 292
Db 230 AEITFGSES-----GKDLSDHQSSSYTLWLDKSTESSVIYVSGTMVLSKKQIEELAR 284
QY 293 GLLAGCRPFLWMIREQND----DGESEEEELSCI--GELKMKGIYVSWCSQLEVL 347
Db 285 ALIEGGRPFLWITDKLNREAKIEGEEETEIKTAGFRHEEYGMIVSQCIEVLRHR 344
QY 348 ALGCFVTHCGWNSAVRSLCGIPVAVPQWFDOTTNAKLTEDAWGTGVRVRMNEGGVDG 407
Db 345 AIGCFVTHCGWNSLESVLVGPVAVPQWSDQPANAKLEEIWKTVGRVNSEGLVER 404
QY 408 CEIERCVEMVMDGDKTKLVRENAIKWKTLARQA 441
Db 405 GEIMRCLEAVNEA--KSVLERENAKKWRLATEA 436

RESULT 9
O23406 PRELIMINARY; PRT; 458 AA.
AC O23406;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE SIMILAR TO INDOLF-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA BEVAN M., STIEKEMA W., MURPHY G., WAMBUUT R., POHL T., TERRY N.,
RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
RA PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DIESTERHOFT A.,
RA JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,
RA SCHUELLER C., CHALWATZIS N.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97339; CAB10333.1; -.
DR MENDEL; 26727; Arath; 2543; 26727.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 458 AA; 52005 MW; 67943A2F CRC32;

Query Match 36.2%; Score 845; DB 10; Length 458;
Best Local Similarity 41.7%; Pred. No. 6e-62;
Matches 194; Conservative 74; Mismatches 129; Indels 58; Gaps 17;

QY 7 LLATFPAGHINPALQFAKRLK--AGTDVTFSTSYAW--RRMANTASAAAGNPPGLDFV 63
Db 15 LFTVFPAGHINPSLELAKRLAGTISGARVTFAASISAYNRRMFSTENV----PETLIFA 70
QY 64 AFSDGYDGLKPG--GDGKR-----YMEKMKARGSEALRNLL-----LNDDVTFVYSH 111
Db 71 TYSQHDGDKSSAYSDDKSDQDATGNFMSEMRRGKETLTLEDNRKQNRPFCTCVYTI 130
QY 112 LFAWAEVARLSHVPTALLWVEPATVLCIYHFYNGVADIEI---DAGSNEIQLPRLPSL 167
Db 131 LLTWAEALAFS-----IFHYFNGVEDAISEMANTPSSSIKLPSPDLL 174
QY 168 EQRSPLTFLPATPERFL--MMKEKLETGEEKAKVLVNTFDALPDALTAI--DRYELI 225
Db 175 TVRDIPSFISSNVYAFLLPAFREQIDSLKEEINPKILNTFQLEPEAMSSVPDNEKIV 234
QY 226 GIGPLISAFLDGEDPSETS YGGDLFEKSENNVEMLSNPKSVVYSGVSRFPKFA 285
Db 235 PVGFLP-----TLTDF-----SSRGEYIWLDTKADSVLVSGTTLAVLSK 278
QY 286 QMEIGIGLLACGRPFLWMIREQNDGEESEEEELSCIG----ELKMKGIYVSWCSQ 341
Db 279 QLVLCALIQSRPFLWITDKSYRNKEDEQKEE--DCISSFREELDEIGWVSWCQDF 337
QY 342 EVLAHPALGCFVTHCGWNSAVRSLCGIPVAVPQWFDOTTNAKLTEDAWGTGVRV--RM 399
Db 338 RVLNHRSGICFVTHCGWNSLTSLSVSGVPVAVPQWNDQMNNAKLLDCWKTVGRVMEKK 397
QY 400 NEGG--GVDGCEIERCVEMVMDGDKTKLVRENAIKWKTLARQA 442
Db 398 EEEGVVVVDSEIRCEIEVME--DKAEFRGNATRWKDLAAEAV 440

RESULT 10
O23401 PRELIMINARY; PRT; 479 AA.
AC O23401;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE INDOLF-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA BEVAN M., STIEKEMA W., MURPHY G., WAMBUUT R., POHL T., TERRY N.,
RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,

```





[illegible]

DE PUTATIVE INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE).  
 GN T20D16.11.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 OC euphyllophytes: Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLUMBIA.  
 RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,  
 RA SKYES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,  
 RA SOMERVILLE C.R., VENTER J.C.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC002391; AAB87119.1; -  
 DR MENDEL: 26261; Arath:3145;26261.  
 DR PROSITE: PS00375; UDPGT; 1.  
 DR PFAM: PF00201; UDPGT; 1.  
 KW Transferase.  
 SQ SEQUENCE 456 AA; 50713 MW; 9A33B1F9 CRC32;  
  
 Query Match 26.6%; Score 621.5; DB 10; Length 456;  
 Best Local Similarity 34.7%; Pred. NO. 1.8e-43;  
 Matches 156; Conservative 86; Mismatches 171; Indels 37; Gaps  
 6 VLLATPFAQGHINPALQFAKRLRLKAGTDVTF-TTSVYAWRMAMTASAAAGNPPGLDFVA 64

[illegible]

```

Db 123 HNISCAILWIOAGAYSVYRYMYKNSPPDLEDL-NQTVLPALPLEVRLDPSMLPS 181
QY 180 TPERFLMKKXLETLDGEEKAVLNTDALEPDALTAI-DREYELIGLPLPSAFDLC 238
Db 182 GGAHFYMLMAEFADCL--RYVKKVLVNSYELESEIESMADLKVPVIGPLV-SPFLIG 238
QY 239 EDPSETSYGGDLFEKSEENNCVWLNKSKSSVYVYVSGSVLRFKPAQMEIEIGKGLLACG 298
Db 239 DGEETLDCKNLDLFCSDCCHEWLDKQARSSVYVYVSGSMLETLENQVETTAKALKNRG 298
QY 299 RFLPMIRKQNDGDEEEBEEELSCIGELKMKG--IVSWCSQLEVLALHPALGCFVTHC 356
Db 299 LPFLWVIR-----PKEKAQNAVVLQEMVKEGQGVLEWSPOEKILSHEAISCFCVTHC 350
QY 357 GNSAVESLSGIPVAVPQWFDQTTNAKLIEDAWGTGVRVNMEGGVGVDG---CEIER 412
Db 351 GNNSTMTVAGVPVAVPWSWDQTDARLLVDVFGVGRMR---NDSVDGELKVVEEVER 407
QY 413 CVEMVMDGDKTLVRENAIKWKTARQAM 442
Db 408 CIEAVTEGPAAND-IRRAELKRVARLAL 436

```

RESULT 15

```

O22183
ID O22183 PRELIMINARY; PRT; 438 AA.
AC O22183;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE PUTATIVE INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
GN T20D16.12
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLOMBIA;
RA ROUNSLAY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC002391; BAB87106.1;
DR MENDEL; 26262; Atg3145;26262.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 438 AA; 48717 MW; A8F76990 CRC32;

```

Query Match 26.3%; Score 613.5; DB 10; Length 438;  
 Best Local Similarity 34.8%; Pred. No. 7.9e-43;  
 Matches 154; Conservative 86; Mismatches 158; Indels 45; Gaps 18;

```

QY 14 QGHINPALQFAKELLKAGTDVTFSTVYAWRRNANTASAAAGNPPGLDFVAESDGY--DD 71
||||| :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 7 QGHINPMLKFAKHLARTNLTHTTTEQARDLLSSTADEP--HRP-VDLAFFSDGLPKDD 63
||||| :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 72 GLKPGGGRYRNMSEKARGSEALRNLLNN--DDVTFVYVYSHLFAWAAEVARLSHVPTAL 129
| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 64 PRPDPTLAK----SLAKDGAKNLKIIIEKRFDCIISVPFT---PWVPAVAHAHNIPCAI 116
||||| :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 130 LWVEPATVLCIYHYF---NGVADETDAGSNEIQLPRLPSLEORSILPTLLPATPERFRL 186
||||| :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 117 LWTQACGAFSVYRYMYKNSPPDLEDL-NQTVLPALPLEVRLDPSMLPSQGANVNT 175
||||| :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 187 MKKELTLDGEEKAKVLYNTDALEPDALTAI-DREYELIGLPLPSAFDGEDPSETS 245
| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 176 LMAEFADCL--KDVKKVLVNSYELESEIESMSDLKPIIPIGPLV-SPFLLGNDSEKTL 232

```

```

QY 246 YGGDLFEKSEENNCVWLNKSKSSVYVYVSGSVLRFKPAQMEIEIGKGLIACGRPFLMWI 305
Db 233 ---DMWK--VDDYCMEMLDKQARSSVYVYVSGSILAKSLENOVETINTALKNRGVPELWVI 287
QY 306 REOKNDGDEEEBEEELSCIGELKMKG--IVSWCSQLEVLALHPALGCFVTHCGNSAVE 363
Db 288 R-----PKEKGENVQLQEMVKEGQGVVTEWQOQEKILSHMAISCFTHCGWNSTIE 339
QY 364 SLSCGIPVAVPQWFDQTTNAKLIEDAWGTGVRVNMEGGVGVDG---CEIERCVEMVMD 419
Db 340 TVVTGVVAVPVTWIDQDPLDARLLVDVFGVGRMK---NDAIDGELKVAEVEKICBAVTE 396
QY 420 GGDKTLVRENAIKWKTARQAM 442
Db 397 GPAAADM-RRRATELKAARSAM 418

```

Search completed: August 1, 2000, 21:31:35  
 Job time: 6064 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:51:30 ; Search time 59.57 seconds  
(without alignments)  
113.997 Million cell updates/sec

Title: US-09-147-955-4  
Perfect score: 2337  
Sequence: 1 MVRRLVLLATPPAQGHINP.....TKLVRENAIKWTKLARQAMG 443

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	707	30.3	471	2	US-08-466-583-2
2	707	30.3	471	4	PCT-US95-07820-2
3	383.5	16.4	131	2	US-08-466-583-5
4	383.5	16.4	131	4	PCT-US95-07820-5
5	375.5	16.1	471	3	US-09-106-464-2
6	368	15.7	488	2	US-08-797-226-2
7	244.5	10.5	131	2	US-08-466-583-6
8	244.5	10.5	131	4	PCT-US95-07820-6
9	219	9.4	63	2	US-08-466-583-8
10	219	9.4	63	4	PCT-US95-07820-8
11	160	6.8	529	4	PCT-US92-00282-7
12	155	6.6	531	4	PCT-US92-00282-6
13	146.5	6.3	534	4	PCT-US92-00282-4
14	145.5	6.2	531	4	PCT-US92-00282-5
15	143	6.1	533	4	PCT-US92-00282-3
16	121.5	5.2	506	5	5180581-2
17	119	5.1	56	2	US-08-466-583-4
18	119	5.1	56	4	PCT-US95-07820-4
19	107	4.6	58	2	US-08-466-583-9
20	107	4.6	58	4	PCT-US95-07820-9
21	104	4.5	74	4	PCT-US92-00282-24
22	103.5	4.4	399	1	US-08-096-623A-20
23	101	4.3	52	2	US-08-466-583-7
24	101	4.3	52	4	PCT-US95-07820-7
25	99	4.2	414	2	US-08-750-524-1
26	90	3.9	1271	1	US-08-095-734-2
27	90	3.9	1271	2	US-08-444-623-2
28	90	3.9	1271	3	US-08-471-869-2

29 90 3.9 1271 4 PCT-US94-08267-2 Sequence 2, Appli  
30 88.5 3.8 421 2 US-08-576-626A-53 Sequence 53, Appl  
31 88.5 3.8 834 1 US-07-977-434-10 Sequence 10, Appl  
32 88.5 3.8 834 1 US-08-073-384C-6 Sequence 6, Appli  
33 88.5 3.8 834 1 US-08-254-359A-6 Sequence 31, Appl  
34 88.5 3.8 834 1 US-08-384-490-31 Sequence 6, Appli  
35 88.5 3.8 834 1 US-08-483-043-6 Sequence 31, Appl  
36 88.5 3.8 834 1 US-08-459-383-31 Sequence 10, Appl  
37 88.5 3.8 834 1 US-08-458-819-10 Sequence 6, Appli  
38 88.5 3.8 834 2 US-08-481-238-6 Sequence 6, Appli  
39 88.5 3.8 834 2 US-08-471-066B-6 Sequence 6, Appli  
40 88.5 3.8 834 2 US-08-484-956-6 Sequence 6, Appli  
41 88.5 3.8 834 2 US-08-757-653-6 Sequence 6, Appli  
42 88.5 3.8 834 2 US-08-599-491-6 Sequence 6, Appli  
43 88.5 3.8 834 2 US-08-756-386-6 Sequence 6, Appli  
44 88.5 3.8 834 2 US-08-823-516-6 Sequence 6, Appli  
45 88.5 3.8 834 3 US-08-682-853A-6 Sequence 6, Appli

## ALIGNMENTS

RESULT 1  
US-08-466-583-2  
; Sequence 2, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szczygowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: and Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,583  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 471 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-466-583-2

Query Match 30.3%; Score 707; DB 2; Length 471;  
Best Local Similarity 37.4%; Pred. No. 5.6e-66;  
Matches 174; Conservative 70; Mismatches 171; Indels 50; Gaps 16;

```
QY 6 VLLATFPAGQHINPALQFAKRLKAGTVDVTFSTVYAWRRMANTASAAAGNPPGLDFVAF 65
Db 5 VLVVFPFGGHNPMVQFAKRLASKGVATTLVTT-----RFIQTADVDA-HPAMVE--AI 57
QY 66 SDGYDDGLKPGDGG-KRYMSEMKARGSEALRNLL-----LNDDVTFVYVSHLFAWAAEVA 120
Db 58 SDGHDEGGFASAGVAEYLEKQAAAASLASLVEARASSADAFTCVVYDSYEDWVLPVA 117
QY 121 RLSHVPTALLWVEPAIVLCIYHFYFNG-----YADEIDAGSNEIQLPR---LPSLE 168
Db 118 RMGLPVPFSTQSCAVSAYVYHFSQGRVAVPPGAAADGSDGAGAAALSEAFGLPEME 177
QY 169 QRSPLTFLLPATPERFRMLMKKLETLDGEEKAKVLVNTFDALPDALTAIDRY-ELIGI 227
Db 178 RSELPSEFVFDHGYPYPTIAMQAIKQFAHAGKDD-WVLFNSFEELETEVLAGLTKYLKARAI 236
QY 228 GPLIP-----SAFLDGEDPSTSYGGDLFEKSEENNCVENLNSPKSVVYVSGSVLR 281
Db 237 GPCVPLPTAGRTAGANG-----RITYGANLVK--PEDACTKWLDTKPDRSVAVVSGSLAS 290
QY 282 FPKAOMEETGKLLACGRPFVLMIREQKNDGDEEEBEEELSCIGELKKMG--KIVSWCS 339
Db 291 LGNAQKEELARGLLAAGKPFVWVR-----ASDEHQVPRYLLAEATATGAAMVVPWCP 343
QY 340 QLEVLHAHPALGCFVTHCGWNSAVESLSCGIPVAVPQWFDQTTNAKLIEDANGTGVVRVM 399
Db 344 QLDVLAHPAVGCFVTHCGWNSLEALSFGVPMVAMALWTDQPTNARNVELANGAGVRRAR 403
QY 400 NEGGV--DGCIEERCVENMVGDKTKLVRENAIKWKTARQAM 442
Db 404 DAGAGVFLRG-EVERCVRAVMDGGEAAASAAKAAAGWRDRARA 447
```

```
RESULT 2
PCT-US95-07820-2
; Sequence 2, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
```

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07820-2

Query Match 30.3%; Score 707; DB 4; Length 471;
Best Local Similarity 37.4%; Pred. No. 5,6e-66;
Matches 174; Conservative 70; Mismatches 171; Indels 50; Gaps 16;

QY 6 VLLATFPAGQHINPALQFAKRLKAGTVDVTFSTVYAWRRMANTASAAAGNPPGLDFVAF 65
Db 5 VLVVFPFGGHNPMVQFAKRLASKGVATTLVTT-----RFIQTADVDA-HPAMVE--AI 57
QY 66 SDGYDDGLKPGDGG-KRYMSEMKARGSEALRNLL-----LNDDVTFVYVSHLFAWAAEVA 120
Db 58 SDGHDEGGFASAGVAEYLEKQAAAASLASLVEARASSADAFTCVVYDSYEDWVLPVA 117
QY 121 RLSHVPTALLWVEPAIVLCIYHFYFNG-----YADEIDAGSNEIQLPR---LPSLE 168
Db 118 RMGLPVPFSTQSCAVSAYVYHFSQGRVAVPPGAAADGSDGAGAAALSEAFGLPEME 177
QY 169 QRSPLTFLLPATPERFRMLMKKLETLDGEEKAKVLVNTFDALPDALTAIDRY-ELIGI 227
Db 178 RSELPSEFVFDHGYPYPTIAMQAIKQFAHAGKDD-WVLFNSFEELETEVLAGLTKYLKARAI 236
QY 228 GPLIP-----SAFLDGEDPSTSYGGDLFEKSEENNCVENLNSPKSVVYVSGSVLR 281
Db 237 GPCVPLPTAGRTAGANG-----RITYGANLVK--PEDACTKWLDTKPDRSVAVVSGSLAS 290
QY 282 FPKAOMEETGKLLACGRPFVLMIREQKNDGDEEEBEEELSCIGELKKMG--KIVSWCS 339
Db 291 LGNAQKEELARGLLAAGKPFVWVR-----ASDEHQVPRYLLAEATATGAAMVVPWCP 343
QY 340 QLEVLHAHPALGCFVTHCGWNSAVESLSCGIPVAVPQWFDQTTNAKLIEDANGTGVVRVM 399
Db 344 QLDVLAHPAVGCFVTHCGWNSLEALSFGVPMVAMALWTDQPTNARNVELANGAGVRRAR 403
QY 400 NEGGV--DGCIEERCVENMVGDKTKLVRENAIKWKTARQAM 442
Db 404 DAGAGVFLRG-EVERCVRAVMDGGEAAASAAKAAAGWRDRARA 447
```

```
RESULT 3
US-08-466-583-5
; Sequence 5, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
```

```
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
PCT-US95-07820-5

Query Match      16.4%; Score 383.5; DB 4; Length 131;
Best Local Similarity 50.7%; Pred. No. 8.6e-33;
Matches 70; Conservative 22; Mismatches 37; Indels 9; Gaps

QY 259 CVEHLNKPSSVVVYSFGSVLRFPPRAQMEIEKGLLLACGRPLWIRBOKNDGDGEEEE 318
Db 1 CTWKLDTPDRSVAIVSYFSGSLAGLSNAQKBELARGLLAAGKPFLLWVR-----ASDEH 53
    | :|||::|| ||||||: || ||: |||| |:|||::||
QY 319 EEELSCLIGELKKMG--KIVSWCSOLEVLAPALPGCFVTHCGWSAVESLSGIPVVPQP 376
Db 5 QVPRYLAEATATGAAMVFWCPQLDVLHPAYVCFFVTHCGWNSTLEALSFGVPWMVAMAL 113
    : :| | | :|||::|||::|:|||::|:|||::|:|||::|:|||::
QY 377 WFDQTTNAKLIEDAWGTG 394
Db 114 WTDPPTNARNVELAWGAG 131
    | |||||: | || |
    | |||||: | || |

RESULT 5
US-09-106-464-2
; Sequence 2, Application US/09106464
; Patent No. 6011145
; GENERAL INFORMATION:
; APPLICANT: Steffens, John C.
; APPLICANT: Ghangas, Gurdev S.
; APPLICANT: Kual, Jian-Ping
; APPLICANT: Eannetta, Nancy
; TITLE OF INVENTION: Chain Length Specific UDP-Glc:Fatty Acid
; TITLE OF INVENTION: Glucosyltransferases
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones, Tullar & Cooper, P.C.
; STREET: P.O. Box 2266 Eads Station
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,464
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,554
; FILING DATE: 13-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Specter, Eric S.
; REGISTRATION NUMBER: 22495
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-415-1500
; TELEX: 703-415-1508
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

US-09-106-464-2

Query Match 16.1%; Score 375.5; DB 3; Length 471;  
Best Local Similarity 27.5%; Pred. No. 4.7e-31;  
Matches 137; Conservative 65; Mismatches 173; Indels 123; Gaps 22;

QY 11 FP--AOGHINPALQAFKRLKAGTDTFTT-----SVYAWRRMAN-----TAS 51  
DB 10 FPMAGHWPFLDMKLVACGCVKATITTPLESVFSKATERKHLGIEIDIRLLKPP 69  
QY 52 AAGNPP-----GLDFVAFSDGDDGLKPGDGKRYMSEMKARGSEALRNLLNDDVTFV 107  
DB 70 AKENDLPEDCERLDLPVSDKLPNFKAAAMKDEFEE-----LIGECRPOCL 117  
QY 108 VYSHLFANAFAEVARLSHVPATLLWVEPATVLCIYHFYNGYADEIDAGSNEIQLPLPSL 167  
DB 118 VSDMFLPMTTDSAAKFSIPRIIVFHGTSYFALCV-----GDSI----- 154  
QY 168 EQKSLP-----TFLLPATPERFLMMKE--KLETLD-----GEEKAK-- 202  
DB 155 -RRNPFKNVSSDTFTFVDPDPHEIRLTRTQLSPFEQSDSETGTMAPMIKAVRESDAKSY 213  
QY 203 -VLVNTFDALPDALTDYR-ELIG-----IGPLIPSAFLDGEDPSETSYGGDLFEKS- 254  
DB 214 GVIFNSFYELSD---YVSHYTKVGVGRKNWAIGPL-----SLCNRD---TEYKAERGRKSS 263  
QY 255 -BENNCVHLNSKPKSSVYVYVSGSVLPPKQAMBEIGKGLLACGRPFLWMIREQND-- 311  
DB 264 IDEHACLKWLDSKSSIVYVCGSTADTTAQMQLAMGLASGODFTWIVRTGNEDWL 323  
QY 312 -DGEIEEEELSCIGELKMKKIV-SWCSQLEVLHAHPALGCFVTHCGWNSAVESLSCI 369  
DB 324 PGFEERTKEK-----GLIRGMAPOVLILDHEAIGAFVTHCGWNSLGLISAGV 373  
QY 370 PVVAVPQWPDQTTNAKLIDANGTVGVY-----RNNEGGVGDGCEIERCVENVMYDGGDKT 424  
DB 374 PMLTVPVFAEQPFNEKLVTEVMRSGAGVSGKQWKRTASEGVKREATAKAIRVM-ASEET 432  
QY 425 KLVRENAIKWTLARQAM 442  
DB 433 EGFRSRAKEYKEMAREAI 450

RESULT 6  
US-08-797-226-2  
; Sequence 2, Application US/08797226  
; Patent No. 5959180  
; GENERAL INFORMATION:  
; APPLICANT: MOEHS, CHARLES P  
; APPLICANT: ALLEN, PAUL V  
; APPLICANT: ROCKHOLD, DAVID R  
; APPLICANT: STAPLETON, ANDREW  
; APPLICANT: GARBARINO, JOAN E  
; APPLICANT: FRIEDMAN, MENDEL  
; APPLICANT: BELKNAP, WILLIAM R  
; TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE  
; TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE  
; TITLE OF INVENTION: GLYCOALKALOIDS IN SOLANACEOUS PLANTS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; STREET: 800 BUCHANAN ST.  
; CITY: ALBANY  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94710  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/797,226  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: PARSONS, NANCY J  
REGISTRATION NUMBER: 40,364  
REFERENCE/DOCKET NUMBER: 0011.97  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 559-5731  
TELEFAX: (510) 559-5777  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 488 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-797-226-2

Query Match 15.7%; Score 368; DB 2; Length 488;  
Best Local Similarity 23.1%; Pred. No. 3.1e-30;  
Matches 117; Conservative 93; Mismatches 172; Indels 124; Gaps 20;

QY 6 VLLATFPAGQHINPALQAFKRLKAGTDTFTTSTVYAWRRMAN----- 49  
DB 13 VLFLPFLSAGHIFPLVNAARLFASRGVKATILTPHNALLFRSTIDDDVRIISGFPISIVT 72  
QY 50 ---ASAAAGNPGCLDFVAFSDGDDGLKPGDGKRY--MSEMKARGSEALRNLLNDDV 104  
DB 73 IKFPAEVLPGEGI-----ESFNSATSPMPHKIFYALSLLQKPMEDKIRELRPD----- 122  
QY 105 TVVYVSHL-FANAFAEVARLSHVPATLLWVEPATVLCIYHFYNGYADEIDAGSNEIQLP- 162  
DB 123 --CIFSDMYFPWTVDIADDELHIPRLYNLSAYMCYSIMH-----NLKVYRPH 167  
QY 163 RLPSPLEQRSPLTFLPATPERFLMMKEKLETL-----DGEKAKYVL 204  
DB 168 KQPNLDESQ--SFVVPGLPDEIKFKLSQLTDLRLKSDQKTVFDELLEQVEDSEERSYGI 225  
QY 205 V-NTFDALPDALTDYR-ELI-----GIGPLIPSAFLDGEDPSETSYGGDLFEK---S 254  
DB 226 VHDFTFYELEP---AYVDYVQKLLKPKCMHFGPL-----SHFASIRSKELIS 269  
QY 255 EENN---CVELNSKPKSSVYVYVSGSVLPPKQAMBEIGKGLLACGRPFLWMIREQND 311  
DB 270 EHNNEIVIDNLNAQPKSVLYVSGSMARFPESQLNEIAQALDASNVFFIFVLR----- 324  
QY 312 DGEIEEEELSCIGELKMKGK-----IVSWCSQLEVLHAHPALGCFVTHCGWNSAVESLSC 367  
DB 325 ---PNEETASWLPVGNLEDKTKGLYIKGWVPQLTMEHSATSGGTGCHGTSVLEAIF 381  
QY 368 GIPVAVPQWPDQTTNAKLE-----DANGTGTVVRNNEGGVGDGCEIERCVC--- 415  
DB 382 GVPMTWPLYADQFYNEKVEYVRGLGKIGKIDVWNEGIEI---TGPVIESAKIREAIEL 438  
QY 416 MYMDGGDKTKLVRENAIKWTLARQA 441  
DB 439 MISNGSEEIINIRVMAKMAQNA 464

RESULT 7  
US-08-466-583-6  
; Sequence 6, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szeszen, Jędrzej B.  
; APPLICANT: Szczygłowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: and Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.



```
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA: US 08/265,427
; APPLICATION NUMBER: 24-JUN-1994
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; US-08-466-583-8

Query Match          9.4%; Score 219; DB 2; Length 63;
Best Local Similarity 65.5%; Pred. No. 5.7e-16;
Matches 36; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

QY 334 IVSWCSQLEVLAPALGCFVTHCGWNSAVESLSGIPVAVVQWFQDTTNKALIE 388
Db 9 VVPWCPQLDVLAPHPAGVCFVTHCGWNSLTLEALSFGVPMVAMALWTDQPTNARNVE 63

RESULT 10
PCT-US95-07820-8
; Sequence 8, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerzen, Jędrzej B.
; APPLICANT: Szczylowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; PCT-US95-07820-8

Query Match          9.4%; Score 219; DB 4; Length 63;
Best Local Similarity 65.5%; Pred. No. 5.7e-16;
Matches 36; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

QY 334 IVSWCSQLEVLAPALGCFVTHCGWNSAVESLSGIPVAVVQWFQDTTNKALIE 388
Db 9 VVPWCPQLDVLAPHPAGVCFVTHCGWNSLTLEALSFGVPMVAMALWTDQPTNARNVE 63

RESULT 11
PCT-US92-00282-7
; Sequence 7, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-00282-7

Query Match          6.8%; Score 160; DB 4; Length 529;
Best Local Similarity 22.4%; Pred. No. 2.9e-08;
Matches 72; Conservative 63; Mismatches 126; Indels 60; Gaps 15;

QY 122 LSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNEIQLP-RLPSEORSLEPFLPAT 180
Db 177 LEHIGQS-----PSPSVYVPRFY-----TKFSDBHTFPQRLANFIANILENYHCL 223
QY 181 PERFRLLMKKEKLETLGDEEKAKVLVNTFDALPDALTAIDRYELI--GIGPLIPS-AFLD 237
Db 224 YSKYEILASDLL-----KRDV---SLPALHONSILWL--RYDFVFEYPRVPMNMFIFG 272
```





## PCT-US92-00282-5

; Sequence 5, Application PC/TUS9200282  
; GENERAL INFORMATION:  
; APPLICANT: OWENS, IDA S.  
; APPLICANT: RITTER, JOSEPH K.  
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION  
; TITLE OF INVENTION: THEREIN.  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
; STREET: 1615 L STREET, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20036-5601

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00282  
; FILING DATE: 19920110  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26581  
; REFERENCE/DOCKET NUMBER: 91532-PT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-861-3000  
; TELEFAX: 202-822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 531 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US92-00282-5

Query Match 6.2%; Score 145.5; DB 4; Length 531;  
Best Local Similarity 24.6%; Pred. No. 9.9e-07;  
Matches 52; Conservative 40; Mismatches 86; Indels 33; Gaps 8;  
  
QY 229 PLIPSAFIDGDPSTSYGGDLPEKSENNVWLNPKSVVYVSPGS-VLRPPKQAQ 286  
DB 265 PVPNNVFVGGINCKRRK---DLSQFE-----AYINASGEHGVVFSLSGSMVSEIPEKK 316  
  
QY 287 MEEIGKGLLACGRPFLWMIREQKNDGDEEEELSCIGELKMKGIYVSCSQALEVLAH 346  
DB 317 AMAIADALGKNPQTVLWRTGTGPSN-----LANNTILVKWLPQNDLLGH 361  
  
QY 347 PALGCFVTHCGNNAVSESCGIPVAVPQWFDQTTNAKLIEDAMGTGVRVRMNEGGYD 406  
DB 362 PMTRAFITHAGSHGVYESICNGVPMVMPFLFGDQMDNAKRMETK-GAGVTLVNLEMTSED 420  
  
QY 407 GCEIERCVEMVMDGGDKTKLVRENAIKKWTLL 437  
DB 421 ---LENALKAVIN-----DKSYKENIMRLSSL 444

## RESULT 15

PCT-US92-00282-3  
; Sequence 3, Application PC/TUS9200282  
; GENERAL INFORMATION:  
; APPLICANT: OWENS, IDA S.  
; APPLICANT: RITTER, JOSEPH K.  
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION  
; TITLE OF INVENTION: THEREIN.  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
; STREET: 1615 L STREET, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00282  
; FILING DATE: 19920110  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26581  
; REFERENCE/DOCKET NUMBER: 91532-PT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-861-3000  
; TELEFAX: 202-822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 533 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US92-00282-3

Query Match 6.1%; Score 143; DB 4; Length 533;  
Best Local Similarity 24.6%; Pred. No. 1.8e-06;  
Matches 47; Conservative 33; Mismatches 77; Indels 34; Gaps 6;  
  
QY 258 NCVE-----WLNKPKSVVYVSPGS-VLRPPKQAQMBEIEIGKGLLACGRPFLWMIR 306  
DB 279 NCLHQNPQSQEPAYINASGEHGVVFSLSGSMVSEIPEKKAMAIADALGKNPQTVLWRYT 338  
  
QY 307 EKNDDGDEEEELSCIGELKMKGIYVSCSQALEVLAHPALGCFVTHCGNNAVSELS 366  
DB 339 GTRPSN-----LANNTILVKWLPQNDLLGHPMTAFITHAGSHGVYESIC 383  
  
QY 367 CGIPVAVPQWFDQTTNAKLIEDAMGTGVRVRMNEGGYDGCCEIERCVEMVMDGGDKTKL 426  
DB 384 NGVPMVMPFLFGDQMDNAKRMETK-GAGVTLVNLEMTSED---LENALKAVIN-----DKS 435  
  
QY 427 VRENAIKKWTLL 437  
DB 436 YKENIMRLSSL 446  
  
Search completed: August 1, 2000, 19:51:31  
Job time: 18753 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:50:19 ; Search time 47.63 Seconds  
(without alignments)  
220.301 Million cell updates/sec

Title: US-09-147-955-4  
Perfect score: 2337  
Sequence: 1 MVRRLVLTAPPAQHINPA.....TKLVRENAIKWTLARQAMG 443

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2337	100.0	443	1 W92948	WO9905287 Seq ID 8
2	2223	95.1	460	1 W92947	WO9905287 Seq ID 7
3	1583.5	67.8	461	1 W92949	WO9905287 Seq ID 9
4	1412.5	60.4	478	1 W92950	WO9905287 Seq ID 1
5	1307.5	55.9	468	1 W92952	WO9905287 Seq ID 1
6	707	30.3	471	1 W03756	Maize UDP-glucose
7	659	28.2	466	1 W09825	UDP-glucose:thiohy
8	384	16.4	433	1 W13406	Solanum melongena
9	378.5	16.2	471	1 W98009	Medium-chain UDP-g
10	374.5	16.0	470	1 W47172	Glucosyl transfera
11	368	15.7	488	1 W64569	Potato solanidine
12	328.5	14.1	471	1 R49245	Anthocyanidin-3-gl
13	324	13.9	473	1 W56451	UDP-glucose:flavon
14	135.5	5.8	534	1 R26154	HUG-Br2. Isolated
15	134	5.7	533	1 R26153	HUG-Br1. Isolated
16	125	5.3	530	1 W47126	uridine diphospho-
17	124	5.3	535	1 W56750	Ecdysteroid UDP-gl
18	121.5	5.2	506	1 R10429	Ecdysteroid UDP-gl
19	110.5	4.7	399	1 R13989	Zeaxanthin glycosy
20	104	4.5	74	1 R30165	UGT1 Exon 4 produc
21	103	4.4	1198	1 W64384	S. cerevisiae L947
22	101	4.3	473	1 W64392	S. cerevisiae L947
23	99	4.2	414	1 R93982	Saccharothrix aero
24	96.5	4.1	421	1 W99388	S. erythraea desosa
25	96	4.1	431	1 Y00181	Enterococcus faeca
26	96	4.1	474	1 W35712	Enterococcus faeca
27	94	4.0	508	1 W35712	Torenia flavonoid
28	91.5	3.9	594	1 R34936	CENP-B. Human cent
29	91.5	3.9	1012	1 W63671	Polypeptide having
30	90	3.9	931	1 Y00872	Original S. tubero
31	89.5	3.8	725	1 W98106	Human elongation f
32	88.5	3.8	421	1 W19735	Sugar biosynthesis
33	88.5	3.8	758	1 R23165	Mutant thermostabl

34	88.5	3.8	789	1 R23164	Mutant thermostabl
35	88.5	3.8	834	1 R13143	T. thermophilus DN
36	88.5	3.8	834	1 R26476	Mutant thermostabl
37	88.5	3.8	834	1 R64274	T. thermophilus DN
38	88.5	3.8	834	1 W79962	Thermus thermophil
39	88.5	3.8	834	1 W80429	DNA polymerase enz
40	88	3.8	906	1 Y07060	Renal cancer assoc
41	86.5	3.7	192	1 Y07089	Renal cancer assoc
42	86.5	3.7	680	1 R23166	Mutant thermostabl
43	86.5	3.7	834	1 R99542	Thermus flavus DNA
44	86.5	3.7	834	1 R96203	DNA polymerase I h
45	86.5	3.7	1891	1 W52844	Anycolatopsis medi

## ALIGNMENTS

RESULT 1

W92948  
ID W92948 standard; Protein; 443 AA.  
AC W92948; 1999 (first entry)  
DT 14-MAY-1999 (first entry)  
DE WO9905287 Seq ID 8.  
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.  
OS Perilla frutescens.  
PN WO9905287-Al.  
PD 04-FEB-1999.  
PF 16-JUL-1998; J03199.  
PR 25-JUL-1997; JP-200571.  
PA (SUNR ) SUNTORY LTD.  
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
PI Yamazaki M.  
DR WPI; 99-142940/12.  
DR N-PSDB; X02827.  
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
PT used to transform plants for improvement of plant coloration  
PS Claim 2; Page 74-76; 89pp; Japanese.  
CC This invention describes the isolation of plant proteins which have  
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used  
CC in the production of plant varieties with improved colour, for example in  
CC the production of flower varieties with desired coloration and cut  
CC flowers harvested from them.  
SQ Sequence 443 AA;

Query Match 100.0%; Score 2337; DB 1; Length 443;  
Best Local Similarity 100.0%; Pred. No. 5.8e-233;  
Matches 443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MVRRRLVLTAPPAQHINPALQFAKRLKACGTDVTFSTVYAWRRMANTASAAAGNPPGL	60
Db	1	MVRRRLVLTAPPAQHINPALQFAKRLKACGTDVTFSTVYAWRRMANTASAAAGNPPGL	60
QY	61	DFVAFSDGYDGLKPGDGKRYMSEMKARGSEALRNLLNNDDVTFVYVSHLFAWAAEVA	120
Db	61	DFVAFSDGYDGLKPGDGKRYMSEMKARGSEALRNLLNNDDVTFVYVSHLFAWAAEVA	120
QY	121	RLSHVPTALLMVEPATVLCIYHFYFNGYADEIDAGSNEIQLPRLPSLEQSRSLPTFLPAT	180
Db	121	RLSHVPTALLMVEPATVLCIYHFYFNGYADEIDAGSNEIQLPRLPSLEQSRSLPTFLPAT	180
QY	181	PERFLRMWKELETDGEEKAKVLNVTDALEPDLTAIDRYELIGIGPLPSAFLDGED	240
Db	181	PERFLRMWKELETDGEEKAKVLNVTDALEPDLTAIDRYELIGIGPLPSAFLDGED	240
QY	241	PSETSYGDLFEKSENNVCNWKPKSSVYVYVSGSVLFPKQAEIEIGKLLACGRP	300
Db	241	PSETSYGDLFEKSENNVCNWKPKSSVYVYVSGSVLFPKQAEIEIGKLLACGRP	300
QY	301	FLWMIREQKNDGDEEEEEEELSCIGELKMGKIVSWCSQLEVLHAHPALGCFVTHCGWNS	360
Db	301	FLWMIREQKNDGDEEEEEEELSCIGELKMGKIVSWCSQLEVLHAHPALGCFVTHCGWNS	360







Plants genetically transformed to interfere with UDP-glucose:thiohydroximate S-glucosyltransferase gene expression - useful for production of rapeseed oil with reduced glucosinolate content

Claim 9: Page 23-25; 35pp; English.

A UDP-glucose:thiohydroximate S-glucosyltransferase (S-GT) (W09825) is encoded by clone pGL9 (766166) amplified from *Brassica napus* cv. Westar cDNA. S-GT is the enzyme responsible for the biosynthesis of glucosinolate. Novel chimeric genes encode an antisense RNA complementary to all or part of an mRNA, the cDNA of which is contained in pGL9. Oilseed rape plants transformed with these chimeric genes have reduced contents of glucosinolates, pref. alkanyl glucosinolates. This allows the prodn. of rapeseed oil with a low glucosinolate content.

Sequence 466 AA;

[illegible]

RESULT	8	
W13406		
ID	W13406	standard; Protein: 433 AA.
AC	W13406;	
DT	10-JUN-1997	(first entry)
DE	Solanum melongena	flavonoid-3-glucosyl transferase.
KW	Eggplant; flavanoid-3-glucosyl;	transferase; glucose; flavonoid;
OS	Pigment; production; seed; plant.	
PN	Solanum melongena.	
PD	J09056385-A.	
PD	04-MAR-1997.	
PF	25-AUG-1995;	238943.
PR	25-AUG-1995;	JP-238943.
PA	(KIRI )	KIRIN BREWERY KK.
DR	WPI:	97-206631/19.
DR	N-PSDB;	T62123.
PT	Solanum melongena	flavonoid-3-glucosyl transferase and DNA encoding
PT	it -	catalyses transfer of glucose to 3-position of flavonoid
PT	pigment in plants	

PS Claim 4; Pages 9-10; 14pp; Japanese.  
CC The present sequence is the Solanum melongena (eggplant)  
CC flavanoid-3-glucosyl transferase (F3GT), which catalyses the  
CC transfer of glucose to the 3-position of flavonoid pigments. The  
CC F3GT DNA is useful for the production of the seed of a plant. A  
CC S. melongena seed was cultured under a red light and a flavanoid  
CC derived. A cDNA library in which the pigment was concentrated was  
CC prepared. A flavanoid glucosidising enzyme gene was cloned and its  
CC structure and mRNA expression pattern analysed. The clone was  
CC expressed in E. coli.  
SQ Sequence 433 AA;

Query Match	16.4%;	Score 384;	DB 1;	Length 433;
Best Local Similarity	25.6%;	Pred. No. 3.5e-31;		
Matches 121;	Conservative 90;	Mismatches 167;	Indels 94;	Gaps

  

QY	6	VLLATFPAGOHINPALQFAKRL---	LKAGTDVTFYFSVYAWRRMANTAAAGNPPGLDF	62
DB	8	IALFAPFGTHAPLLTLVQKISPFLLSPSTIFSPFWT-----	SSNSISFSKVPNQEN	60
QY	63	VAFSDGVDDGLKPGGDKRYMSEMKARGSDALR-----	NLLLN-----DDVTFV----	108
DB	61	IKIYNW-DGVKRGND-----	TPFGLPAIKLFQSTLLISKITEEAEEETGVAFSCI	111
QY	109	YSHLEFAWA--AEARLSHVPTALLVPEPATVLCIYHFNGYADETDAG-SNEIQLPLRP	165	
DB	112	FSDAFLMCFVLKLPKKNAAPGVAYTGGGCSLAVH-----	YDLRSNKETSLKIPGFS	166
QY	166	S-LEQRSPLTFLPATPERPRIMMKKLE-----	TLGEEKAKVLVNTFDALP	214
DB	167	STLSINDIP-----PE-----	VTAELEGPMSSMLYNALNLHRKADAVVLSFQELDRD	215
QY	215	ALTAID-----RYELIGTGPLIPSAFLDGEDPSETSYGGDLFEKSENNCVWLNSPKKS	270	
DB	216	PLINKDLQKNLQKVFNGIPVLQAS-----	SRKLDESGCIIQLWDKQEKKS	259
QY	271	VYVYFGSVLRFPAQMEETGKGLLAGRRPFLMWIREQKNDGDEEEEEEISCIGELKK	330	
DB	260	VVYLSFGVTATLPNEIGTISAEALTKTTFPIWSLR-----	NNGVKNLPKGFLE--RTKE	312
QY	331	MGIKTVSCSOLEVLHAPALGCFVYTHCGWNSAVESLSGPIPVAVPQWFQDTTKAKLIEDA	390	
DB	313	FGKIVSWAPQLEIAHAKSVGFVYTHCGWNSILGEGISFGVPMICRPFPGDQKLSRWAVESV	372	
QY	391	WGTGVRVYRMNKGSGVDGCETERCVERVMYDGGDKTKLVRENAIKWTKLARQAM	442	
DB	373	WEIGLOI-----EGGFTKSGIISALDAPFFN--EKGKILRENVEGLKEKALEAV	420	

RESULT 9

W98009

ID W98009 standard; Protein; 471 AA.

DT 21-JUN-1999 (first entry)

DE Medium-chain UDP-glucose:fatty acid glucosyltransferase.

KW UDP-glucose:fatty acid glucosyltransferase; tomato;

KW 1-O-acyl-a-glucose; acyl donor; esterification.

OS Lycopersicon pennellii

OS Lycopersicon esculentum.

PN W90909144-AA.

PD 25-FEB-1999.

PF 10-AUG-1998; IB1369.

PR 30-JUN-1998; US-106464.

PR 13-AUG-1997; US-055554.

PA (CORR ) CORNELL RES FOUND INC.

PI Eanneatta NT, Giangas GS, Kuai J, Steffens JC;

DR WPI: 99-190155/16.

DR N-PSDB: X24873.

PT New chain length specific uridine diphosphate-glucose:fatty acid

PT glucosyltransferases - useful for preparing straight or branched,

PT (un)saturated 1-O-acyl-a-glucoses and 1-O-acyl-a-D-glucoses

PS Example 3; Page 36-38; 42pp; English.

CC The present sequence is a UDP-glucose:fatty acid glucosyltransferase  
 CC that transfers glucose to medium-chain length fatty acids to form  
 CC 1-O-acyl-a-glucoses. The 1-O-acyl-a-glucoses act as acyl donors  
 CC in the esterification of glucose and in the further esterification  
 CC of partially acylated glucose in reactions catalysed by glucose  
 CC acyltransferases. They also act as acyl donors in the  
 CC esterification of sucrose and in the further esterification of  
 CC partially acylated sucrose in reactions catalysed by sucrose  
 CC acyltransferases. A cDNA clone (see X24873) encoding the enzyme  
 CC was isolated by immunoscreening a leaf trichome cDNA library  
 CC prepared from an FI population of a cross between wild tomato  
 CC Lycopersicon pennellii (LA 716) and cultivated tomato  
 CC Lycopersicon esculentum cv. New Yorker. The invention provides  
 CC a chain length-specific UDP-glucose:fatty acid glucosyltransferases  
 CC and a method for their purification. The enzymes have specificity  
 CC for transferring glucose to short, medium or long chain length  
 CC fatty acids. Methods for preparing 1-O-acyl-a-D-glucoses, where  
 CC the acyl group comprises 3-5, 6-13 or 14-22 C atoms, are claimed.  
 SQ Sequence 471 AA;

Query Match 16.2%; Score 378.5; DB 1; Length 471;  
 Best Local Similarity 28.1%; Pred. No. 1.5e-30;  
 Matches 136; Conservative 69; Mismatches 184; Indels 95; Gaps 21;

QY 11 FP--AOGHINPALQFAKRLKAGTDVTF--SVYAWRRMAN-----TAS 51  
 DB 10 FPMAGCHMPTLDMAKLVACRGVKATITTPNESVFSKAIERNKHLGIEIDIRLLKPP 69  
 QY 52 AAGNPP-----GLDFVAFSDGDDGLKPGDGKRYMSEMKARGSEALRLLNDDVTF 107  
 DB 70 AKENDLPEDCERLDLPDSDKLPNFKAAAMKDEFE-----LIGECRPDCL 117  
 QY 108 VYSHLFAWAAEVARLSHVPTALLWVEPAIVLCIYHFY-----FNGYADEIDA-----GSN 157  
 DB 118 VSDMFPLPWTDSAAKFSIPRIVFHGTSYFALCVGHSIRRNKPKFNYSSTETFWVPDPFH 177  
 QY 158 EIQLPR--LPSLEQSLPTFLPATPERFLMMKLETLGDEKAKVLVNTFDALPEPA 215  
 DB 178 EIRLTQLSPPESQDEETGMAP-----MIKAVRES--DAKSGVIFNSFYELSD- 226  
 QY 216 LTAIDRY-ELIG-----IGLIPSAFLDGEDPSETSYGGDLFEKS--EENNCVWELNPK 267  
 DB 227 --VVEHYTKVVGKKNWAIGPL-----SLCNRD--IEYKAERGRKSSIDEHACLKWLDSKK 277  
 QY 268 KSSVYVSGSVLRFPKQAEIEIGKLLACGRPFLMWIREOKND---DGEEREEBELSC 324  
 DB 278 SSSIVYVCGSTADFTTAQMOELAMGLEASGQDFIWIIRTGNEDWLPGEFTEKREK--- 334  
 QY 325 IGEKMKMKIV-SWCSQLEVLHAPALGCFVTHCGWNSAVESLSGIPVAVPQWFDQTN 383  
 DB 335 -----GLIIRGWAPOVLILDHEAIGAFVTHCGWNSTLEGISAGVPMLTWPVFAQFFN 387  
 QY 384 AKLIEDAWGTGVRV-----RMNEGGVGDCEIERCERVMYMDGDKTKLVRENAIKWTLA 438  
 DB 388 EKLVTVMRSGAGVGSKQWKRTASEGVKREAIKAKIKRYM--ASEETEGRSRAKEYKEMA 446  
 QY 439 RQAM 442  
 DB 447 REAI 450

RESULT 10

W47172  
 ID W47172 standard; Protein; 470 AA.  
 AC W47172;  
 DT 08-JUN-1998 (first entry)  
 DE Glucosyl transferase (Gtase) protein encoded by TW1 gene.  
 KW Glucosyl transferase; Gtase; TW1; tomato; signalling pathway;  
 KW salicylic acid; jasmonic acid; ethylene; wound inducible gene;  
 KW plant defence protein; plant response; tobacco; rice.  
 OS Lycopersicon sp.  
 PN W09745546-A1.

PD 04-DEC-1997.  
 PF 30-MAY-1997; G01473.  
 PR 31-MAY-1996; GB-011420.  
 PA (UYO-) UNIV YORK.  
 PI Bowles DJ, Calvert CM, O'Donnell PJ, Roberts MR;  
 DR WPI; 98-032653/03.  
 DR N-PSDB; V17054.  
 PT Tomato wound inducible (TW1) gene encoding glucosyl transferase -  
 PT useful to develop products that alter signalling pathways in plants  
 PT by altering of salicylic acid, jasmonic acid or ethylene  
 PS Claim 2; Fig 3; 52pp; English.  
 CC This is a glucosyl transferase (Gtase) protein encoded by a wound  
 CC inducible gene (TW1) isolated from wounded tomatoes. The TW1 gene  
 CC encodes this Gtase from amino acid position 5. The TW1 gene can be  
 CC used to identify homologue Gtase encoding genes isolated from tobacco  
 CC and rice. A microbial host can be transfected or transformed with a  
 CC vector containing the Gtase encoding nucleic acids. The products can be  
 CC used to interfere with Gtase and therefore alter signalling pathways in  
 CC plants, specifically tobacco, rice or tomato plants by altering levels  
 CC of salicylic acid, jasmonic acid or ethylene. This can induce the  
 CC production of plant defence proteins such as pathogenesis-related (PR)  
 CC and proteinase inhibitor (PIN) proteins which regulate plant development  
 CC (plant growth, reproduction and senescence) and improve plant response to  
 CC pathogens.  
 SQ Sequence 470 AA;

Query Match 16.0%; Score 374.5; DB 1; Length 470;  
 Best Local Similarity 27.4%; Pred. No. 3.8e-30;  
 Matches 136; Conservative 64; Mismatches 177; Indels 119; Gaps 21;

QY 11 FP--AOGHINPALQFAKRLKAGTDVTF--SVYAWRRMAN-----TAS 51  
 DB 9 FPDAGCHMPTLDMANVACRGVKATITTPNESVFSKAIERNKHLGIEIDIRLLKPP 68  
 QY 52 AAGNPP-----GLDFVAFSDGDDGLKPGDGKRYMSEMKARGSEALRLLNDDVTF 107  
 DB 69 AKENDLPEDCERLDLPDSDKLPNFKAAAMKDEFE-----LIGECRPDCL 116  
 QY 108 VYSHLFAWAAEVARLSHVPTALLWVEPAIVLCIYHFYFNGYADEIDAGSNEIQLPRLPSL 167  
 DB 117 VSDMFPLPWTDSAAKFSIPRIVFHGTSYFALCV-----GDTI----- 153  
 QY 168 EQRSLP-----FLLPATPERFLMMKE--KLETLD-----GEEKAK-- 202  
 DB 154 -RNKPKFNVSSTETFWVPDLPHETRLTQLSPPESQDEETGMAPMIKAVRESDAKSY 212  
 QY 203 -VLVNTFDALPEDALTAIDRY-ELIG-----IGLIPSAFLDGEDPSETSYGGDLFEKSE 255  
 DB 213 GVIFNSFYELSD---VVEHYTKVVGKKNWAIGPL-SLCNRDIEDKAERGRKSSI---D 264  
 QY 256 ENNCVWELNPKSKSSVYVSGSVLRFPKQAEIEIGKLLACGRPFLMWIREOKND---D 312  
 DB 265 EHACLKWLDSKSSSIVYVCGSTADFTTAQMOELAMGLEASGQDFIWIIRTGNEDWLP 324  
 QY 313 GEEREEBELSCIGELKMKMKIV-SWCSQLEVLHAPALGCFVTHCGWNSAVESLSGIPV 371  
 DB 325 GFERTKEK-----GLIIRGWAPOSVILDHAEIAGAFVTHCGWNSTLEGISAGVPM 374  
 QY 372 VAVPQWFDQTNNAKLIEDAWGTGVRV-----RMNEGGVGDCEIERCERVMYMDGDKTKL 426  
 DB 375 VTPVFAEQFFNEKLVTEVMRSGAGVGSKQWKRTASEGVKREAIKAKIKRYM--ASEETE 433  
 QY 427 VRENAIKWTKLARQAM 442  
 DB 434 FRSRAKEYKEMAREAI 449

RESULT 11

W64569  
 ID W64569 standard; Protein; 488 AA.  
 AC W64569;  
 DT 07-DEC-1998 (first entry)



DE Potato solanidine UDP-glucose glucosyltransferase.  
 KW Solanidine UDP-glucose glucosyltransferase; SGT; potato; tomato;  
 KW glycoalkaloid; solasodine; solanaceous plant; transgenic plant.  
 OS Solanum tuberosum cv. Lemhi Russet.  
 FH Key Location/Qualifiers  
 FT Region 12..46  
 FT /note= "conserved domain"  
 FT Region 110..143  
 FT /note= "putative steroid binding region"  
 FT Region 351..401  
 FT /note= "UDP-glucose binding region"  
 PN W09834471-A1.  
 PD 13-AUG-1998.  
 PF 30-JAN-1998; U01864.  
 PR 07-FEB-1997; US-797226.  
 PA (USDA ) US SEC OF AGRIC.  
 PI Allen PV, Belknap W, Friedman M, Moehs CP, Rockhold DR,  
 PI Stapleton A;  
 DR WPI; 98-446828/38.  
 DR N-PSDB; V49609.  
 PT New DNA encoding solanidine UDP-glucose glucosyl-transferase and  
 PT related vectors - and transgenic plants, used to reduce contents of  
 PT steroidal glyco-alkaloid(s) in plants, specifically potato  
 PS Example 1; Fig 2; 5app; English.  
 CC This is the amino acid sequence of potato cv. Lemhi Russett  
 CC solanidine UDP-glucose glucosyltransferase (SGT). It was deduced  
 CC from an SGT cDNA sequence (see V49609). SGT is involved in the  
 CC biosynthesis of steroidal glycoalkaloids in solanaceous plants,  
 CC catalysing the UDP-glucose dependent conversion of the aglycone  
 CC solanidine to gamma-chaconine. The invention relates to DNA  
 CC sequences which encode SGT, and their use, particularly use of an  
 CC antisense DNA construct to inhibit SGT activity and glycoalkaloid  
 CC levels in solanaceous plants. Transgenic plants are claimed,  
 CC particularly tomato and potato, that have reduced contents of toxic  
 CC steroidal glycoalkaloids. The inactivation of glycoalkaloid  
 CC biosynthetic pathways is beneficial to reduce or eliminate  
 CC glycoalkaloid biosynthesis during storage and shipping.  
 SQ Sequence 488 AA;

Query Match 15.7%; Score 368; DB 1; Length 488;  
 Best Local Similarity 23.1%; Pred. No. 1.9e-29;  
 Matches 117; Conservative 93; Mismatches 172; Indels 124; Gaps 20;

QY 6 VLLATPFAQGHINPALQFAKRLKAGTDVFTFTSYAVRRMANT----- 49  
 DB 13 VLFPLFSAGHPIPLVNAARLFASRGVATILTTTHNALLFSTIDDDVRISGPFISIVT 72  
 QY 50 ----ASAAAGNPPGLDFVAFSDGYDGLKPGDGKRY--MSENKARGSEALRNLLNDDV 104  
 DB 73 IKFPAEVLGPEGI-----ESFNATSPMPHKIFYALSLLQKPMEDKIRELRPD---- 122  
 QY 105 TFVYVSHL-FAWAEVARLSHVPTALLWVEPATVLCIYHFYNGVADEIDAGSNEIQLP- 162  
 DB 123 --CIFSDMYFPTVDIADLHPIRLYILNLSAYMCYSIMH-----NLKVYRPH 167  
 QY 163 RLPSEQRSLPTFLPATPERELMMKELETL-----DGEKAKVL 204  
 DB 168 KQPNLDESQ--SFVVPGLPDEIKFKLSQTLDDLRKSDQKTVFDELLEQVEDSEERSVGI 225  
 QY 205 V-NTDALEPDAITADRYELI-----GIGPLIPSAFLDGEDPSETSYGDLDFEK--S 254  
 DB 226 VHDTFVELEP---AVVDYQKLKPKCKWHFGPL-----SHFASKIRSKELIS 269  
 QY 255 ENNN---CVELNLSKPKSVVYVSGSVLRFPFKAQMEIEIGKLLACGRPFLLWIREQKND 311  
 DB 270 EHNNNEIVIDWLNQPKSVLYVSGSMARFPESQLNETAQALDASNVPPFIFVLR----- 324  
 QY 312 DGEEREEELSCIGELKMKGK-----IVSWCSQLEVLAPALGCFVTHCGWNSAVESLSC 367  
 DB 325 ----PHEETASWLPVGNLEDKTKKGLYIKGWVPQPTIMEHSATGGFWTHCGTNSVLEAITF 381  
 QY 368 GIPVAVVQWFDQTTNAKLIE-----DAWGTGVRVRNMEGGVDGCBIERCVE--- 415

DB 382 GVPMTWPLADQFYNEKVVVEVRLGKIGIKIDWNEGIEI---TGPVIESAKIREAIERL 438  
 QY 416 MYMDGCDKTKLVRENAIKWKTLARQA 441  
 DB 439 MISNGSEIIRDRVMAWSKMAQNA 464

RESULT 12  
 R49245  
 ID R49245 standard; Protein; 471 AA.  
 AC R49245;  
 DT 08-AUG-1994 (first entry)  
 DE Anthocyanidin-3-glucoside rhamnosyltransferase (3RT).  
 KW Anthocyanidin-3-glucoside rhamnosyltransferase;  
 KW glucosyltransferase; inflorescence; flowering plants;  
 KW transgenic plant; Petunia hybrida.  
 OS Petunia hybrida.  
 PN W094003591-A.  
 PD 17-FEB-1994.  
 PF 30-JUL-1993; AU0387.  
 PR 30-JUL-1992; AU-003846.  
 PA (ITFL-) INT FLOWER DEV PTY LTD.  
 PI Brugillera F, Holton TA;  
 DR WPI; 94-065680/08.  
 DR N-PSDB; Q56241.  
 PT Nucleic acid encoding glucosyltransferase enzymes - used for  
 PT producing transgenic plants with altered inflorescence properties  
 PT including modified petal colours  
 PS Claim 24; Page 52-55; 7opp; English.  
 CC The DNA encoding the anthocyanidin-3-glucoside rhamnosyltransferase  
 CC is used in the construction of a vector which can then used  
 CC to transform plants. The transgenic plants are then capable of  
 CC expressing the anthocyanidin-3-glucoside rhamnosyltransferase which  
 CC confers altered inflorescence properties including modified petal  
 CC colours.  
 SQ Sequence 471 AA;

Query Match 14.1%; Score 328.5; DB 1; Length 471;  
 Best Local Similarity 26.6%; Pred. No. 2.2e-25;  
 Matches 124; Conservative 70; Mismatches 200; Indels 73; Gaps 16;

QY 6 VLLATPFAQGHINPALQFAKRLKAGTDVFTFTSYAVRRMANTASAAAGNPPGLDFVAF 65  
 DB 12 VYVFFFAEGHISPFVQLANKLSSYGVKVSFFFTASGNASRVKSMLSA----PTTHIVPL 67  
 QY 66 SDGYDDGLKPGDGKRYMSEKARGSEALR-----NLLNDDVTFFVYVSHLFAW 115  
 DB 68 TLPHVEGLPPGAES---TAELTPASAEALLKVALDLMQPOIKTLLSHLKPHEVLFDPQEW 124  
 QY 116 AAEVARLSHVPTALLWVEPATVLCIYHFYNGVADEIDAGSNEIQLPRLPSLEQRSLPTF 175  
 DB 125 LPKMANGLGIKTVYVSVVVALSTAF-----LTCPARVLEPKPKYPSLEDMKKPPL 173  
 QY 176 LIPAT-----PERFLMMK-----EKLET-LDGEKAKVLYNTFEDALEPDAL 216  
 DB 174 GPPQSVTSVTRFEARDFLYVFKSPHNGPTLYDRQSGLRG--CSAILAKTCSQMEGPI 231  
 QY 217 TAID---RYELIGIPLIPSAFLDGEDPSETSYGDLPEKSENNCVELNLSKPKSVVY 273  
 DB 232 KYVEAQFNKPVFLIGPVVP-----DPP-----SGKLEEK-----WATWLNKPEGTVIY 275  
 QY 274 VSGSVLRFPFKAQMEIEIGKLLACGRPFLLWIREQKN-DGEEEREEELSCIGELKMKG 332  
 DB 276 CSFGSETFTJDDQVKELALGLEQTGLPFFLVLPANVDVSAELNRPALPEGFLEVRKDKG 335  
 QY 333 KIVS-WCSQLEVLAPALGCFVTHCGWNSAVESLSCGIPWVAVPQWFDQTTNAKLIEDAW 391  
 DB 336 IHHSWVQOOHTLAHSSVCYCHAGFSVIEALYNDCCQVMLPQKGDGDIILNAKLVSMD 395  
 QY 392 GTGVRVRNMEGGVDGCE-IERCVENVMVDGDKT--KLVRENAIKWK 435

```

Db 396 EAGVEINRDEDDGYGKEDIKEAVEKVMVDVEKEPGKLIRENQKWK 442

RESULT 13
W56451
ID W56451 standard; Protein: 473 AA.
AC W56451;
DE 06-AUG-1998 (first entry)
KW UDP-glucose:flavonoid 3,5-O-glucosyl transferase amino acid sequence.
KW UDP-glucose:flavonoid 3,5-O-glucosyl transferase; gentian; rose plant;
KW blue flower.
OS Unidentified.
PN J10113184-A.
PF 15-OCT-1996; 272253.
PR 15-OCT-1996; JP-272253.
PA (TOLG ) TOKYO GAS CO LTD.
DR WPI; 98-315473/28.
DR N-PSDB; V23108.
PT Gentian flavonoid glucosyl transferase gene - used to produce
PT recombinant rose plants with blue petals.
PS Claim 1; Pages 6-7; 10pp; Japanese.
CC The present sequence represents a UDP-glucose:flavonoid 3,5-O-glucosyl
CC transferase enzyme. The nucleic acid sequence was amplified from
CC RNA which had been isolated from gentian (sic) petals. The gene is
CC used to transform rose plants to produce blue flowers.
SQ Sequence 473 AA;

Query Match 13.9%; Score 324; DB 1; Length 473;
Best Local Similarity 24.4%; Pred. No. 6.4e-25;
Matches 120; Conservative 84; Mismatches 179; Indels 108; Gaps 18;

QY 1 MVRVRLATFPAGQHINPALQAKLLKAGTDVTF-----TSYVAMRRMANTAS 51
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MKKAELVIPPFGISHLGSVVELAKLAERNEHLSIVIIKEPNDTKVSNLLKSLSTAS 60
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 52 -----AAGNPPGLDFVAFSDGVDGLKPGDGKRYMSEMKARGSEALRNL-LLNN 101
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 RIKVIELKQETISIEGTGLFI-----QKFIESHKAQVRDFLAGVSNCS 104
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 102 DQVTFVYVYSHLFAWAEVARLSHVPTALLWVEPATVLCIYHFYNGYADE-----IDAGSN 157
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 105 VELSGVVIDMFCSTSMIDVANEFEVPSVFTSSAAMLGLW-FHFQSLRDNGFYVDIKDS 163
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 158 EIOLPRLPSEQSLPTF--LLPATPERFLMMKKEKLETLDG-----EEKAKVL 204
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 ETVL-----SIPAFQNLVP-----YGLPWFIFNTEDGCDASLDIGKRFRETKGII 209
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 205 VNTFDALPDALTAIDRYELI---GIGLPIPSAFLDGEDPSETSYGGDLFEKSENNCV 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 210 INTFLESHAIESTLSTDETIIPPVTVGPIL-----GPK-----GSSIESLETEKIL 256
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 261 EWLNKPKSSVYVSGSLRFPKAQMEETGKGLACGRPFLMWIREQ-----KNDGEE 316
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 257 KWLDMQPEKSWVFLFCGLGHGFEAQVKEIAYALEGSGHRLFLSLRKPPLPLGKFEQGEY 316
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 317 EEEELSCIGELKK---MGMKIVSWCSQLEVLAPALGCCFTHCGWNSAVESLSCGIPVA 373
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 317 ENLEEVLPGEFLERTANTGVMGAPQATVLSHSAVGGFVSHCGWNSLTSIWFQGPVAT 376
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 374 VPQWFQTTNAKLIEDAWGTVGVVRNM---EGGVDGCEIERC-----VEVMYMDGDKT 424
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 377 WPLFAEQONNAFVELVELGLAVEVKMDYKDKYNPDADVLVRADVTEEKIKILMD----- 431
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 KLVRENAIKWK 435
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 432 ---PENGIRKK 439
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
R26154
ID R26154 standard; Protein: 534 AA.

```

```

AC R26154;
DE 27-JAN-1993 (first entry)
DE HUG-Br2
KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW monoglucuronide; diglucuronide.
OS Homo sapiens.
FH Key
FT region
FT 12..22
FT /note= "putative membrane-insertion signal"
FT 492..508
FT /note= "putative membrane-anchoring peptide"
FT modified_site 348
FT /note= "predicted Asn-linked glycosylation site"
FT misc_difference 282..285
FT /note= "residues encoded by TGCCAACGGGAAG 1"
PN W09212987-A.
PD 06-AUG-1992.
PF 10-JAN-1992; U00282.
PR 10-JAN-1991; US-639453.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Owens IS, Ritter JK;
DR WPI; 92-284593/34.
DR N-PSDB; Q27369.
PT Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT types I and II
PS Disclosure; Fig 9A-I; 99pp; English.
CC Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC been isolated. They are referred to as HUGBr1 (Q27369) and HUGBr2
CC (Q27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC upon expression individually in COS-1 cells, encode isoforms that
CC catalyse the formation of the two bilirubin monoglucuronides and
CC the diglucuronide.
CC The cDNAs contain identical 3' ends (1469 bp in length) to each
CC other and to that of the human phenol transferase cDNA, HUGP1
CC (Harding et al., Proc. Natl. Acad. Sci. USA 85:8281 (1988)).
CC In contrast, they have unique 5' ends.
CC Sequence 534 AA;

Query Match 5.8%; Score 135.5; DB 1; Length 534;
Best Local Similarity 24.2%; Pred. No. 2.3e-05;
Matches 51; Conservative 38; Mismatches 89; Indels 33; Gaps 8;

QY 229 PLTPS-AFLDGEDPSETSYGGDLFEKSENNCVEMLNKPKSSVYVSGS-VLREPKAQ 286
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 PIMPNVFIGG---INCANGKPLSQEPE-----AYINASGEHGIIVFSLSEMSVEIPEKK 319
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 287 MEEIGKGLLACGRPFLMWIREQKNDGDEEEEEEELSCIGELKKMKIKIVSWCSQLEVLAH 346
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 AMAIADALGKIPQTVLWRYTGTRPSN-----LANNTILVKWLQNDLLGH 364
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 347 PALGCFVTHCGWNSAVESLSCGIPVAVPQWFDQTTNAKLIEDAWGTVGVVRNNEGVD 406
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 365 PMTRAFTHAGSHGVESICNGVPMVMPPLFGDQMDNAKRMETK-GAGVTLNVLMTSED 423
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 407 GCIEERCVENMDGGDKTKLVRENAIKWKTL 437
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 ---LENAQKAVIN-----DKSYKENIMRLSSL 447
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
R26153
ID R26153 standard; Protein: 533 AA.
AC R26153;
DE 27-JAN-1993 (first entry)
DE HUG-Br1.
KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW monoglucuronide; diglucuronide.
OS Homo sapiens.
FH Key
FT region
FT /note= "putative membrane-insertion signal"

```



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 14:38:12 ; Search time 3616.02 Seconds  
(without alignments)  
1797.580 Million cell updates/sec

Title: US-09-147-955-3  
Perfect score: 1474

Sequence: 1 accaaccaaacacaaattt.....acttaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	gb_est1:*
21:	gb_est2:*
22:	gb_est3:*
23:	gb_est4:*
24:	gb_est5:*
25:	gb_est6:*
26:	gb_est7:*
27:	gb_est8:*
28:	gb_est9:*
29:	gb_est10:*
30:	gb_est11:*
31:	gb_est12:*
32:	gb_est13:*
33:	gb_est14:*
34:	gb_est15:*
35:	gb_est16:*
36:	gb_est17:*
37:	gb_est18:*
38:	gb_est19:*
39:	gb_est20:*
40:	gb_est21:*
41:	gb_est22:*
42:	gb_est23:*
43:	gb_est24:*

44:	gb_est25:*
45:	gb_est26:*
46:	gb_est27:*
47:	gb_est28:*
48:	gb_est29:*
49:	gb_est30:*
50:	gb_est31:*
51:	gb_est32:*
52:	em_est20:*
53:	em_est21:*
54:	em_est22:*
55:	em_est23:*
56:	em_est24:*
57:	em_est25:*
58:	em_est26:*
59:	gb_est33:*
60:	gb_est34:*
61:	gb_est35:*
62:	gb_est36:*
63:	gb_est37:*
64:	gb_est38:*
65:	em_est27:*
66:	em_est28:*
67:	em_est29:*
68:	em_est30:*
69:	gb_est39:*
70:	gb_est40:*
71:	gb_est41:*
72:	gb_est42:*
73:	gb_est43:*
74:	gb_est44:*
75:	em_est31:*
76:	em_est32:*
77:	em_est33:*
78:	em_est34:*
79:	gb_est45:*
80:	gb_est46:*
81:	gb_est47:*
82:	em_est35:*
83:	em_est36:*
84:	em_est37:*
85:	gb_est48:*
86:	gb_est49:*
87:	gb_est50:*
88:	gb_est51:*
89:	gb_est52:*
90:	gb_est53:*
91:	gb_est54:*
92:	gb_est55:*
93:	gb_gss1:*
94:	gb_gss2:*
95:	gb_gss3:*
96:	gb_gss4:*
97:	em_gss1:*
98:	em_gss2:*
99:	em_gss3:*
100:	em_gss4:*
101:	gb_gss5:*
102:	gb_gss6:*
103:	gb_gss7:*
104:	gb_gss8:*
105:	gb_gss9:*
106:	em_gss5:*
107:	em_gss6:*
108:	em_gss7:*
109:	em_gss8:*
110:	em_gss9:*
111:	em_gss10:*
112:	em_gss11:*
113:	gb_gss10:*
114:	gb_gss11:*
115:	em_gss12:*
116:	gb_gss12:*



QY	799	taactcggtgaggttggaactcgaagccgaatacttcggtggtgtatgtgtcgtttg	858
Db	58	TGACTACATGGAATGGTTAAACACAAAGCCATAATCATCAATTCGTTTATATCTCATTTGG	117
QY	859	gagcgttttgaggtttccaaagcacaataatggaagagattgggaagggctattagcctg	918
Db	118	GAGTCTATTGAATTTATCAAGAAACCAAAAGGAGGAGATTGCAAAAGGGTTGATAGAGAT	177
QY	919	cgaagggccctttttatggatgatacgaagaacagaagaatgacgacgacggaagaaga	978
Db	178	CCAAAGGCCATTCTATGGGTATAAGAGATCA-----AGNAGAAAGAAA	222
QY	979	agaagaagaagttgagttgcatgctgggaattgaaaaaatggggaataatagtgctg	1038
Db	223	AGAAAGAGAGAAATTAAGTTGTCATGATGGAATTAGAAGAACGAAGGGAAATAGTACCATT	282
QY	1039	gtactcagttgaggttctggcgaccctgcgttgggattgttctgacgcattatg	1098
Db	283	GTGTTACAACTTGAAGTCTTGACACATCCATCTTTAGGATGTTTCTCGCACTGTGG	342
QY	1099	gtggaactcggcttgagagcttgagttgcgggattcccggtggtggtggtgcgcagtg	1158
Db	343	ATGGAATTCGACTCTAGAAAGCCATGCTCAGGTGTACCTGCTGTGTCATTCCTCATTTG	402
QY	1159	gtttgatacagcagcgaatgcgaagctgattgagatgctgctgggggacaggggtgagagt	1218
Db	403	GACAGATCAAGGGGACTAATGCCAAAGTTGATTGAAGATGTTTGGAAAGACAGGTGTCAGAT	462
QY	1219	gagaataaatgaaggggtgggtggtgattgagatgagatagagaaggtgtgtggagatggt	1278
Db	463	GAGAGTCAGTGAAGATGGTGTGGTTGAGATGTGTAATAAAAGGGTCATAGAAATGT	522
QY	1279	gatggtgggggtg	1292
Db	523	AATGGATGGTGGAG	536
RESULT	3		
AI897621			
LOCUS	546 bp	mrna	EST
DEFINITION	EST267064 tomato ovary, TAMU Lycopersicon esculentum cDNA clone		27-JUL-1999
ACCESSION	U018821, mRNA sequence.		
VERSION	AI897621		
KEYWORDS	AI897621.1 GI:5603523		
SOURCE	EST.		
ORGANISM	tomato.		
REFERENCE	Lycopersicon esculentum		
AUTHORS	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.		
TITLE	1 (bases 1 to 546)		
JOURNAL	Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E., Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and Giovannoni, J.		
COMMENT	Generation of ESTs from tomato carpel tissue Unpublished (1999) On Apr 7, 1998 this sequence version replaced gi:3035329. Contact: David Frisch Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfrisch@CLEMSON.EDU 5 prime sequence. Location/Qualifiers 1..546 /organism="Lycopersicon esculentum" /cultivar="TA496" /db_xref="taxon:4081" /clone="cLED28K21"		
FEATURES			
source			





Qy	978	aagaagaagaggttgagttgcattgggggaataaaaaaatggggaataatggtcgt	1037
Db	343	AAGAAGAAGAGAAATTAAGTTCATGATGGAATTAGAGACAAGGGAATAATGATACCAT	402
Qy	1038	ggtagtcgagttgagagttctggcgcaacctcggttgggaatggtttcgtagcgaattgtg	1097
Db	403	GGTGTTCAACAACCTTGAAGTCTCGACATCATCTTTAAAGATGTTTGTCTCGCACTGTG	462
Qy	1098	ggtggaactcggtcgtg 1113	
Db	463	GATGGAATTCGACTCT 478	
RESULT	6		
AW349414/c			
LOCUS	AW349414	747 bp	mrna
DEFINITION	GN210007A20D2R Gm-r1021 Glycine max cDNA 3'	EST	01-FEB-2000
ACCESSION	AW349414		
VERSION	AW349414.1	GI:6847124	
KEYWORDS	EST.		
SOURCE	soybean.		
ORGANISM	Glycine max		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.		
AUTHORS	1 (bases 1 to 747) Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corvella, V., Erpelting, J., Raph, C., Shoop, E., Pardini, J., Liu, L. and Lewin, H. A Functional Genomics Program for Soybean (NSF 9872565) Unpublished (1999)		
TITLE	On Jul 9, 1999 this sequence version replaced gi:5434916.		
JOURNAL	Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)		
COMMENT	Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics University of Illinois Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA Tel: (217) 244-6147 Fax: (217) 333-4582 Email: l-vodkin@uiuc.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.		
FEATURES	Location/Qualifiers		
source	1..747 /organism="Glycine max" /cultivar="Williams" /db_xref="taxon:3847" /clone_lib="Gm-r1021" /tissue_type="root" /lab_host="XL10-Gold" /note="vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; Library pBluescript II is a sequence-driven, reracloned set of the original library Gm-cl004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #20401) was used to synthesize the cDNA. The Gm-cl004 library was constructed by Dr. Paul Keim & Virginia H. Corvella, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.keim@nau.edu, virginia.corvella@nau.edu. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html. Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois,		

http://www.life.uiuc.edu/biotech/keck.html.\*

BASE COUNT	206 a	235 c	97 g	187 t	22 others
ORIGIN					
Query Match	14.1%	Score	207.2;	DB 71;	Length 747;
Best Local Similarity	58.7%	Pred.	No. 7.7e-46;		
Matches	419;	Conservative	0;	Mismatches	253; Indels 42; Gaps 3;
QY	719	attccctcgcctcttcttgacgagcgaagacatccctcgaacgtcttctacgagcgccgatctt	778		
Db	733	ATTCNNNNNGCCTTCNNNGCCTTAAGATCCCGGATACCTNNNTATCGCGGTGATNNN	674		
QY	779	tctcgaataatcgagagagaataactcgtgaggtgttgaactcgaagcgcgaatcttcg	838		
Db	673	NCCNNTGCTTC-----AAATGATACGTTGATGTTGGCTCAGCATCAGAGCTGATATCT	620		
QY	839	gtggtgtatgtcgtttgggagcgttttgaggtttccaaagcacaataatgaagagatt	898		
Db	619	GTGGTTTATGTTTCAATTTGTTACCCCTTGCTGTGTTGGCTGATAGACAGATGAAGGAGCTT	560		
QY	899	gggaaagggctattagcctcggaagcgcctttttatgtagatatacgaacaagaagaat	958		
Db	559	GCACGCGCCTGCTAGATTCCGGATATCTTCTTGTGGCTCATTAGAGATATG-----	504		
QY	959	gacgagcgcgaagaagaagaagaagaagattgattgattgggaattgaaaaa	1018		
Db	505	-----CANGAATAGAGATTAATGTCAGAGAGAACTGGAGCAG	467		
QY	1019	atggggaaaaatagtgctggtgctgcagttgaggtttcgcgcacccctcggttggga	1078		
Db	466	AGGGGTAAGATTGTGAATGGTGTCTCAGTGGAGGTTCTGCGCATGGTTCCTTGGGT	407		
QY	1079	tgttcgtgacgcattgtggttggaactcggctgtgagagcttgattgctgggattccg	1138		
Db	406	TGTTTGTAAACGATTTGGATGGAATTCGACTATGGAATGTTGGTTCGGGGTTCCT	347		
QY	1139	gtggtgctgctgctgagtcggtttgtagcagcagcgaatgcgaagctgattgaggtgctg	1198		
Db	346	ATGGTGCGCTTCCCGCAGTGGACAGACCAAGGACCAATGCAAGATGGTGCAAGATGTG	287		
QY	1199	tgggggacaggggtgagagtgagaatgaa-----tgaaggggggtgggttgatgga	1249		
Db	286	TGGAAGACGGGGTGAGGTGGATGATGAAGGTGAATGTTGAGGAAGGGATAGTGGAGCA	227		
QY	1250	tgtgagatagaaggtgtgtgagatggtgtagatgaggtgggtgacaaagccaaactagtg	1309		
Db	226	GAGGAGATTAGGAAGTGTGTTGGATGTGCTATGGGAGTGGAGGAAGAACAGCAATTC	167		
QY	1310	agagaaaaatgccatcaaatggaagactttggccagacagccatgggataggatgatct	1369		
Db	166	AGAAGGAATGCTGATTAATGGAATGCTTGGCCAGGGAGGCCGTCACCGAAGGAGGCTCT	107		
QY	1370	tcactcaacaatcaaacgcttcttcgtcgaagttgcacacttttaattctgt	1423		
Db	106	TGCGATAGTAATATGAGGACTTTCTCCATGATGTTGCAAAATTTGGACATGAT	53		
RESULT	7				
LOCUS	AI487571	520 bp	mRNA	EST	29-JUN-1999
DEFINITION	EST245893 tomato ovary, TAMU Lycopersicon esculentum cDNA clone				
ACCESSION	AI487571				
VERSION	AI487571.1	GI:4382942			
KEYWORDS	EST.				
SOURCE	tomato.				
ORGANISM	Lycopersicon esculentum				
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				
AUTHORS	Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;				
	Solanaceae; Solanum; Lycopersicon.				
	1 (bases 1 to 520)				
	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,				

BASE COUNT	194 a	76 c	107 g	143 t
ORIGIN				
Query Match	13.6%	Score	201.2;	DB 40;
Best Local Similarity	67.4%	Pred.	No. 3.1e-44;	Length 520;
Matches	322;	Conservative	0;	Mismatches 138; Indels 18; Gaps 2;
QY	591	tgaaggagaagctggaaactttagacggtgaagagagcggaagctattggtgaacact	650	
Db	57	TCAAGAGCAACTAGACACATAGATGGTGAGAGAAATCCAAAGGTACTTGTGAATACAT	116	
QY	651	ttagtcggttggagcccgatgcactcacgctattgtaggttagttagtcggtgattcg	710	
Db	117	TTCATGCTAGAGCTAGAGCCACTCAAAGCTATTCAAATAACAAATTTAATTGGAAATG	176	
QY	711	ggccggttagtccctccgctctttagacgaggaagatccctccgaaacgtcttcggcg	770	
Db	177	GACCATGTGATCTCTTCATCATCTTGGGTGGAAAGATTCATTGGTAATCTTCATTTGGTG	236	
QY	771	gcgactcttcgaaaaatcgagagagaataaactcgcgtgagtggttgaactcgaagcga	830	
Db	237	GTGATCTTTTCAAAGTCAATG---ATGACTACATGGAATGTTAAACACAAAGCCTA	293	
QY	831	aattctcgggtggtgtatgtctgttggagcggttggaggtttccaaagggcacaatgg	890	
Db	294	AATCATCAATTTGTATATCTCATTTGGGAGTCTATTGAATTTATCAAGAAACCAAAAGG	353	
QY	891	aagagattgggaaaggctattagctcgcggaaggccctttttatgtagatagagaac	950	
Db	354	AGGAGATTGCAAAAGGGTGTGATAGATCCAAAGGCCATCTTATGGGTAAATAAGAGATC	413	
QY	951	agaagaatgcgcgcgcgaagagaagaagaagaagcttagttagttagtcattggggaaat	1010	
Db	414	A-----AGAAAGAAAGAAAGAGAGAGAAATTAAGTTGCATGATGGAAAT	458	
QY	1011	tgaaaaaaatgggaaaaatagtgtcgtgctgcagttggaggttcggtgcgcccc	1068	
Db	459	TAGAGAAGCAAGGGAATAATAGTACCATTGTTTCACAACCTTGAAGTCCCTGACACATCC	516	
RESULT	8			
LOCUS	AW442098	649 bp	mRNA	EST
DEFINITION	EST311494 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA			
AUTHORS	AW442098	649 bp	mRNA	EST
	EST311494	tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA		

Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.		
Generation of ESTs from tomato carpel tissue Unpublished (1999)		
On Apr 7, 1998 this sequence version replaced g1:3035667.		
Contact: David Frisch		
Clemson University Genomics Institute		
Clemson University		
100 Jordan Hall, Clemson, SC 29634, USA		
Tel: 864 656 4366		
Fax: 864 656 4293		
Email: dfrisch@CLEMSON.EDU.		
Location/Qualifiers		
1..520		
/organism="Lycopersicon esculentum"		
/cultivar="TA496"		
/db_xref="taxon:4081"		
/clone="cLED10D21"		
/clone_lib="tomato ovary, TAMU"		
/tissue_type="carpel"		
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"		
/lab_host="XLI-Blue MRF"		
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2: XhoI; cLED - Tomato Carpel EST Library. Oligodt-primed and directionally cloned cDNA in vector Lambda ZAP II with 5', and 3' ends located at the EcoRI and XhoI sites, respectively."		
BASE COUNT 194 a 76 c 107 g 143 t		
ORIGIN		
Query Match 13.6%; Score 201.2; DB 40; Length 520;		
Best Local Similarity 67.4%; Pred. No. 3.1e-44;		
Matches 322; Conservative 0; Mismatches 138; Indels 18; Gaps 2;		
QY	591 tgaagagagagctgaaacttttagacggtgagagagcggaagatttggtgaacacgt	650
Db	57 TCAAGAGCAACTAGACACATTAGATGGTGCAAGAAAATCCAAAGGTACTTGTGAATACAT	116
QY	651 ttgatacgttgagacccgcatgcaactcacggctattgataggtgatggtgagatcg	710
Db	117 TTGATGATTTAGAGCTAGAGCCACTCAAGCTATTGAAAAATACAAATTTAAATTTGGAATG	176
QY	711 ggcggttgatccctcgcctcttgagcgcgaagatccctccgaacgtctctacg	770
Db	177 GACCATGATTCCTTCATCATCTTCTGGGTGGAAGAAATTCATTGGAAATCTTCATTTGGTG	236
QY	771 gcgacttttcgaaaaatcgagagagataaactcgtgagtggttgaactcgaagccga	830
Db	237 GTGATCTTTTCAAAAGTCAATG---ATGACTACATGGAATGTTAAACACAAGCCTA	293
QY	831 aactctcgttggtgtatgtcgttttgagagcgtttttgaggtttcccaagggcacaatcg	890
Db	294 AATCATCAATTTGTTATATCTCATCTATTGGGAGTCTATTGAAATTTATCAAGAAGCAAAAGG	353
QY	891 aagagattgggaaaggcgtattagcctcggaagggccctttttatgtagatatacgaac	950
Db	354 AGGAGATTGCAAAAGGTTGATAGAGATCCAAAGGCCATCTCTTATGGGTAAAGAGATC	413
QY	951 agagaatgacgacgagcgaagagagagagagagagaggttgagttgcatggggaat	1010
Db	414 A-----AGAGAAGAAAAAGAGAAGAGAATAATTAAGTTGCATGATGGAAT	458
QY	1011 tgaataaaatgggaaataatgctgctgtgtcgcagttggaggttcttcgagccacc	1068
Db	459 TAGAAGAAGCAAGGAAAAATAGTACCATGCTGTTTCAAACTTGAAGTCTCTGACACATCC	516
RESULT	8	
AW442098	649 bp	mRNA
LOCUS	EST311494	tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA
DEFINITION	14-FEB-2000	





RESULT 12			
AI779099			
LOCUS	AI779099	517 bp	mRNA
			EST
			29-JUN-1999

```

Db 509 A 509

RESULT 13
AW651280
LOCUS EST329734 tomato germinating seedlings, TAMU Lycopersicon
DEFINITION esculentum cDNA clone cLE116K15 5', mRNA sequence.
ACCESSION AW651280
VERSION AW651280.1 GI:7412518
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 658)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and
Tanksley,S.D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
On Jan 6, 2000 this sequence version replaced gi:5677182.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
source
Location/Qualifiers
1..658
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLE116K15"
/clone_lib="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stage="7 days post imbibition"
/notes="Vector: pBluescript SK(-); Site_1: EcorI; Site_2:
XhoI; 7 days post imbibition on water-agar. Mixed stage
whole germinating seedlings from seed coat emergence up
to two centimeters in seeds not showing obvious signs of
germination were discarded."
BASE COUNT 208 a 126 c 124 g 200 t
ORIGIN
Query Match 11.2%; Score 165.4; DB 79; Length 658;
Best Local Similarity 59.0%; Pred. No. 1.8e-34;
Matches 384; Conservative 0; Mismatches 221; Indels 46; Gaps 4;

QY 266 aagcgctacatgccagatgaaagccgcgctccgagccttaagaacctctctc 325
Db 8 AACGTTACATGTCAGAGATAGAACTCTGGTCCCAACATGAGGATGTCATTGTG 67

QY 326 aacaacgagac-----gtcactttcgtcgtctactccacacctctttgcattg 373
Db 68 AAGAGTTCAGACGAGGACGCTGCTGTCACGTCCTCGTCTACACTCTTCTGCTACCTTGG 127

QY 374 gcgcgcgaggtggcggtttgtccacggtccgacgcgcctctctctggtcgagccgcgc 433
Db 128 GCAGCTGAGGTAGCGCGGTGAACCTCCATATATCCATCCGCGCTTACTATGAATTCACACGCA 187

QY 434 accgtctgtgatataccactctctcaacggtctacgacgacgacgacgcgcgt 493
Db 188 ACTGTGCTAGACATATACACTACTTACTTCAATGGCTATGAGGATGAATGAGTGACT 247

QY 494 tccaatga-----aatcagctccctcgcgtctccatccctcgagcgcgc 538

```

```

Db 248 ACCTCAAAATGATCCAAATTTGGAGTATCCAAATTTGCCAAGGCTTCCATTACTAAAGCCAA 307
QY 539 agtcttcgagcgtttctcgtcgtcgacgcccggagagattccggttgatga----- 590
Db 308 GATCTTCCATCTTTTATTTAGTTTCATCTAGCTCAAAAGATGATGAGTATAGTTTGTCTCT 367
QY 591 -----tgaaggagaaagctggaaccttttagcgggtgaagagagcgcgaagatttgt 642
Db 368 ACCAACATTTCAAGAGCAACTAGACACATTTAGTGGTGAAGAAATCCAAAGGTACTTGT 427
QY 643 gaacacgtttgatgcgtttggagcccgatgcactccggtctattgatgatgatgattgat 702
Db 428 GAATACATTTGATGATTTAGCTAGAGTAGAGCCACTCAAAGCTATTGAAAAATACAATTAAT 487
QY 703 cgggatacggcgctgtgattccctcgcctctcttgagcggcgaagatccctcgcgaacgctc 762
Db 488 TGAATTTGGACCATGATTCCTTCATTCATTTCTGGTGGGAAAGATTCATTTGGAATTTTC 547
QY 763 ttacgcggcgctctctttcgaaaaatcggagaggaataaactcgtgaggtggtgaactc 822
Db 548 ATTGGTGGTGATCTTTTCAAAGTCAATG---ATGACTACATGGAATGGTAAACAC 604
QY 823 gaagccgaatacttcggtggtgatgtcgtttgtgggagcgttttgaggtt 873
Db 605 AAAGCCTAAATCATCAATTTGTTTATATCTCATTTAGGAGTCTATTGAATTT 655

RESULT 14
AW034633
LOCUS EST278317 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION cLEC33F2, mRNA sequence.
ACCESSION AW034633
VERSION AW034633.1 GI:5893389
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 612)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S.,
Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
On Jul 7, 1999 this sequence version replaced gi:5405910.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
source
Location/Qualifiers
1..612
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC33F2"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/notes="Vector: pBluescript SK(-); Site_1: EcorI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
BASE COUNT 166 a 136 c 133 g 177 t

```



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:50:31 ; Search time 61.65 Seconds  
(without alignments)  
517.335 Million cell upd

Title: US-09-147-955-2  
 Perfect score: 2421  
 Sequence: 1 MVRRVLLATTPAQGHINP.....GEGSSLKLNLAFTLHOVAR 460

Scoring table: BLOSUM62

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters:	225878
--	--------

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database :
SPREMBL_12:*
1:  sp_archaea:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mmc:*
8:  sp_organelle:*
9:  sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	2421	100.0	460	10	Q9ZR27	perilla fru	
2	2233	91.8	443	10	Q9ZR26	perilla fru	
3	1646.5	68.0	461	10	Q9ZR25	verbena x h	
4	1053.5	43.5	467	10	Q93709	nicotiana t	
5	1038	42.9	456	10	Q23270	arabidopsis	
6	995	41.1	519	10	Q9ZVY2	arabidopsis	
7	971	40.1	474	10	Q04930	arabidopsis	
8	949.5	39.2	455	10	Q9ZVY5	arabidopsis	
9	888	36.7	458	10	Q23406	arabidopsis	
10	716.5	29.6	449	10	Q22822	arabidopsis	
11	702	29.0	479	10	Q23401	arabidopsis	
12	687.5	28.4	449	10	Q22820	arabidopsis	
13	675.5	27.9	460	10	Q48676	arabidopsis	
14	663.5	27.4	456	10	Q22182	arabidopsis	
15	658.5	27.2	484	10	Q23400	arabidopsis	
16	653.5	27.0	438	10	Q22183	arabidopsis	
17	628.5	26.0	475	10	Q23402	arabidopsis	
18	604.5	25.0	453	10	Q22186	arabidopsis	
19	580.5	24.0	481	10	Q9ZVJ3	arabidopsis	

## ALIGNMENTS

```

RESULT 1
Q92R27 PRELIMINARY; PRT; 460 AA.
ID Q92R27
AC Q92R27;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYSNIN 5-O-GLUCOSYLTRANSFERASE.
GN PF3R4.
OS Perilla frutescens..
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
core eudicots; Asteridae; euasterids I; Lamiales; Lamiaceae; Perilla.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. SHIKUN; TISSUE-LEAF;
RX MEDLINE; 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.,"
RL J. Biol. Chem. 274:7405-7411(1999).
RD EMBL; AB013596; BAA36421.1; -.
RW PROSITE; PS00375; UDPGT; 1.
DR Transferase.
KW SEQUENCE 460 AA; 50974 MW; FE7CCF22 CRC32;
SQ
Query Match 100.0%; Score 2421; DB 10; Length 460;
Best Local Similarity 100.0%; Pred. No. 6.1e-190;
Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVRRRVLLATFPAGGHINPALQFAKRLKAGTDTFTFTSVYAWRRMANTASAAAGNPPGL 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 MVRRRVLLATFPAGGHINPALQFAKRLKAGTDTFTFTSVYAWRRMANTASAAAGNPPGL 60
QY 61 DFVAFSDGYDGLKPCGDGGRKYSEMKGARSEALRNLLLNHNDVTFVYSHLFAWAAEVA 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 DFVAFSDGYDGLKPCGDGGRKYSEMKGARSEALRNLLLNHNDVTFVYSHLFAWAAEVA 120
QY 121 RESQVPSALLWEPATVLCITVYFFNGYADEIDAGSDEIOLPRUPPLEORSLPFTLLPET 180

```

```

Db 121 RESQVPSALLWVEPATVLCIYYFYFNGYADEIDAGSDEIQLPRLPLEQRSLPTFLLPET 180
QY 181 PERFRLLMMKEKLETLGDEEKAKVLVNTFDALPDALTAIDRYELIGTGLPISAFLDGGD 240
Db 181 PERFRLLMMKEKLETLGDEEKAKVLVNTFDALPDALTAIDRYELIGTGLPISAFLDGGD 240
QY 241 PSTSYGGDLFEKSENNVCVWLDTPKSSVYVSGSVLRFPPKAQMEIEIGKGLLACGRP 300
Db 241 PSTSYGGDLFEKSENNVCVWLDTPKSSVYVSGSVLRFPPKAQMEIEIGKGLLACGRP 300
QY 301 FLWMIREQKNDGDEEBEELSCIGELKMGKIVSWCSQLEVLALHPALGCFVTHCGWNSAV 360
Db 301 FLWMIREQKNDGDEEBEELSCIGELKMGKIVSWCSQLEVLALHPALGCFVTHCGWNSAV 360
QY 361 ESLSGCVPVVAVPQWFDQTTNAKLIEDAGTGVVRMNEGGVYDGSIEIERCVEMVMDGGE 420
Db 361 ESLSGCVPVVAVPQWFDQTTNAKLIEDAGTGVVRMNEGGVYDGSIEIERCVEMVMDGGE 420
QY 421 KSKLVRENAIKWKTILAREAMGDSLSKLNALFLHVARA 460
Db 421 KSKLVRENAIKWKTILAREAMGDSLSKLNALFLHVARA 460

```

## RESULT 2

```

Q9ZR26 ID Q9ZR26 PRELIMINARY; PRT; 443 AA.
AC Q9ZR26;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE
GN PF3R6
OS Perilla frutescens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Lamiales; Perilla.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-CV, SHIKUN; TISSUE-LEAF;
RX MEDLINE; 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL; AB013597; BAA36422.1; -.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 443 AA; 49110 MW; D87B5486 CRC32;

```

Query Match 91.8%; Score 2223; DB 10; Length 443;  
 Best Local Similarity 95.7%; Pred. No. 8.9e-174;  
 Matches 424; Conservative 9; Mismatches 8; Indels 2; Gaps 1;

```

QY 1 MVRRLVLLATFPAGGHINPALQFAKRLKAGTQVTFSTSYAWRRMANTASAAAGNPGL 60
Db 1 MVRRLVLLATFPAGGHINPALQFAKRLKAGTQVTFSTSYAWRRMANTASAAAGNPGL 60
QY 61 DFVAFSDGYDGLKPCGDGKRYKSEMKARGSEALRNLLNNHNDVTFVYVSHLFAWAAEVA 120
Db 61 DFVAFSDGYDGLKPCGDGKRYKSEMKARGSEALRNLLNNHNDVTFVYVSHLFAWAAEVA 120
QY 121 RESQVPSALLWVEPATVLCIYYFYFNGYADEIDAGSDEIQLPRLPLEQRSLPTFLLPET 180
Db 121 RLSHVPALLWVEPATVLCIYYHYFYFNGYADEIDAGSNEIQLPRLPSLEQRSLPTFLLPAT 180
QY 181 PERFRLLMMKEKLETLGDEEKAKVLVNTFDALPDALTAIDRYELIGTGLPISAFLDGGD 240
Db 181 PERFRLLMMKEKLETLGDEEKAKVLVNTFDALPDALTAIDRYELIGTGLPISAFLDGGD 240

```

```

QY 241 PSTSYGGDLFEKSENNVCVWLDTPKSSVYVSGSVLRFPPKAQMEIEIGKGLLACGRP 300
Db 241 PSTSYGGDLFEKSENNVCVWLDTPKSSVYVSGSVLRFPPKAQMEIEIGKGLLACGRP 300
QY 301 FLWMIREQKNDG--EBEELSCIGELKMGKIVSWCSQLEVLALHPALGCFVTHCGWNS 358
Db 301 FLWMIREQKNDGDEEBEELSCIGELKMGKIVSWCSQLEVLALHPALGCFVTHCGWNS 360
QY 359 AVESLSGCVPVVAVPQWFDQTTNAKLIEDAGTGVVRMNEGGVYDGSIEIERCVEMVMDG 418
Db 361 AVESLSGCVPVVAVPQWFDQTTNAKLIEDAGTGVVRMNEGGVYDGSIEIERCVEMVMDG 420
QY 419 GEKSLVRENAIKWKTILAREAMG 441
Db 421 GDKTKLVRENAIKWKTILARQAMG 443

```

## RESULT 3

```

Q9ZR25 ID Q9ZR25 PRELIMINARY; PRT; 461 AA.
AC Q9ZR25;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE.
GN HGT8.
OS Verbena x hybrida.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Verbenaceae;
OC Verbena.
[1]
RN SEQUENCE FROM N.A.
RC TISSUE-PETAL;
RX MEDLINE; 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL; AB013598; BAA36423.1; -.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 461 AA; 51347 MW; 6DCBC3E6 CRC32;

```

Query Match 68.0%; Score 1646.5; DB 10; Length 461;  
 Best Local Similarity 70.0%; Pred. No. 1.3e-126;  
 Matches 326; Conservative 49; Mismatches 74; Indels 17; Gaps 7;

```

QY 1 MVRRLVLLATFPAGGHINPALQFAKRLKAGTQVTFSTSYAWRRMANTASAAAGNPGL 60
Db 1 MSRRVLLATFPAGGHINPALQFAKRLANADIQVTFSTSYAWRRMRT---AAGSGLI 57
QY 61 DFVAFSDGYDGLKPCGDGKRYKSEMKARGSEALRN--LLNNHND---VTFVYVSHLFA 114
Db 58 NFVSFSDGYDGLQPGDGGKRYKSEMKARGSLKALSDTLAANNVDQKSKITFVYVSHLFA 117
QY 115 WAAEVAESQVPSALLWVEPATVLCIYYFYFNGYADEIDAGSDEIQLPRLPLEQRSLP 173
Db 118 WAAKVAEFHLRSALLWIEPATVLDIFFYFYFNGYDEIDAGSDAIHLPGGLVLAORDLP 177
QY 174 TFLLPETPERFRLLMMKEKLETLGDEEKAKVLVNTFDALPDALTAIDRYELIGTGLPIS 233
Db 178 SFLLPSTHERFRSLMKLETLGEEKPKVLVNSFDALPDALKAIKDYEMIAIGLPIPS 237
QY 234 AFLDGGDSETSGLDLPFK--SENNVCVWLDTPKSSVYVSGSVLRFPPKAQMEIEIGK 292
Db 238 AFLDGGDPSRSGFDGDFEKGSDDDCLEWLSNPRSSVYVSGSVVNTTKSQMEETAR 297
QY 293 GLLACGRPFLLMMIREQKNDGDEEBEELSCIGELKMGKIVSWCSQLEVLALHPALGCFVT 352

```

```

Db 298 GLLDCGRFLVWVRV-----NEGEVLISCMEELKRVGKIVSWCSQLEVLTHPSIGCFVT 352
QY 353 HCGNSAVESLSCGVPVAVVQWFDQDTNNAKLIEDAGCTGVRVRMNEGGG-VDSSEIERC 411
    ||||| :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 353 HCGNSLIESISFGVPVAVVQWFDQDTNNAKLIEDAGCTGVRVRMNEGGG-VVDSGEIIRC 412
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
QY 412 VEMVMDGGEKSLVRENAIKWTKTLAREAMGDGSSKLNNAFLHOV 457
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 413 IEEVMDGGEKSLVRENAIKWTKTLAREAMGDGSSKLNNAFLHOV 458
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :

RESULT 4
P93709 PRELIMINARY; PRT; 467 AA.
AC P93709
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE GLUCOSYL TRANSFERASE.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC Nicotiana.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BY-2; TISSUE-SUSPENSION CULTURE;
RC KOTIMA H., HASHIZUME K., IMANISHI S., NAKAMURA K.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB000623; BAA19155.1; -.
DR MENDEL; 9421; Nicta; 1525; 9421.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 467 AA; 52989 MW; 3726EB31 CRC32;

Query Match 43.58; Score 1053.5; DB 10; Length 467;
Best Local Similarity 45.38; Pred. No. 3.8e-78;
Matches 209; Conservative 83; Mismatches 150; Indels 19; Gaps 7;

QY 6 VLLATFPAGQHINPALQFAKRLKAGTDVTFSTSVYAWRRMANTASAAAGNPPGDFVAF 65
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 VLIATFPQGHINFLQFSKLLINLVKVTLSLSLAFNRKLNPKI-----EGLTAPP 64

QY 66 SDGVDGLKPCGDKR-YMSEMKARGSEALRNLL-----LNNHDTVYVYSHLFAWAEVA 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 SDGYDGNFGKSFDDYHLFNSAIKSGSEFANLIKSAKAKNGYPTFRVITYTILMDWAGSA 124

QY 121 RESOVPSALLWVEPATVLCIYFYFNGYAD---EIDAGSDEIQIPLRPLPLEORSPTFL 177
    : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 KKLHPSITLWQIPATVFDIYIYFTFANFTKKNYDSQDIELPGLPSLSSSDFPSFVF 184

QY 178 PETPERFLM-MKEKLETLDGEEKAKVLNTPDALTDRYELIGIGLPSAF 235
    : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185 DDVNSDWAVESIKRQIILNSENPRILNTPDALELNAVLKNTWVGIGLIPSSF 244

QY 236 LDGSDPSETS YGGDLFKSENNCNVEMWDTKPKSSVYVSGSVLRFPKQAMEIGKGLL 295
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 245 LDEKRDKNFADMIIE--SENNYMEWLDARANKSVIYIAPGSVAEISSQWMEISQGLL 302

QY 296 AGRPFLWMIREQKNDGDEEEELSCIGELKMGKIVSWCSQLEVLHAHPALGCFVTHCG 355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 303 KCGRPFLWVIRETLN--GEKPEELTKCKLEKIGRVRWCSQMEVLKHSVSGCFLTHCG 360

QY 356 WNSAVESLSCGVPVAVVQWFDQDTNNAKLIEDAGCTGVRVRMNEGGGVDGSEIERCEMV 415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 WNSTLESASGVPVACPIWNDQICNAKLIDQVWKIGVRVANKEGIIKRDFOKCEIIV 420

QY 416 MDGGEKSLVRENAIKWTKTLAREAMGDGSSKLNNAFLHQ 456
    : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 MGDAEGEELKNAQKWKDLAKESTKENSNNVNLKAYNE 461

```

```

RESULT 5
O23270 PRELIMINARY; PRT; 456 AA.
AC O23270
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE GLUCOSYL TRANSFERASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98121113.
RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,
RA BERGKAMP R., DIRKSE W., VAN STAVEREN M., STIEKEMA W., DROST L.,
RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFFANELLI P., WEDLER H.,
RA WEDLER E., WAMBUTT R., WEITZENEGGER T., POHL T.M., TERRIN N.,
RA GLENN J., WILLARROEL R., DE CLERCK R., VAN MONTAGU M., LECHARNY A.,
RA AUBORG S., GY I., KREIS M., LAO N., KAVANAGH T., HEMPEL S., KOTTER P.,
RA ENTION K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,
RA SILVEY M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,
RA VOUKELATOU E., MILLONI D., HATZOPOULOS P., PIRAVANDI E., OBERMAIER B.,
RA HUBERT H., DUESTERHOFT A., MOORES T., JONES J.D.G., ENEVA T.,
RA PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,
RA DELSENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,
RA SCHUELLER C., CHALWATZIS N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana";
RL Nature 391:485-488(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 297335; CAB10189.1; -.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 456 AA; 50702 MW; EAE0D380 CRC32;

Query Match 42.98; Score 1038; DB 10; Length 456;
Best Local Similarity 48.28; Pred. No. 6.7e-77;
Matches 225; Conservative 72; Mismatches 134; Indels 36; Gaps 12;

QY 3 RRRVLLATFPAGQHINPALQFAKRLKAGTDVTFSTSVYAWRRMANTASAAAGNPP--G 59
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11 RPHVLLVTFPAGQHINPALQANLRIHHGATVTVSTAVSAHRM-----GEPPSTKG 62

QY 60 LDFVAFSDGYDDGLKPCGDKRMYMSEMKARGSEALRNLLNNHD-----VTFVYVSHL 112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 LSFATFTDGFDDGLKSPEDQKIYHSELKRCGSLNLRDIIKANLDAITTEPTITGVYVSVL 122

QY 113 FAWAAEVARESQVPSALLWVEPATVLCIYFYFNGYADEIDAGSDEIQLRPLPLEORS 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 VPMVSTVAREPHLPTLWIEPATVLDIYIYFNTSYKHLFDV--EPIKLPKLPLITGTD 180

QY 172 LPTLLPETPERFLM-MKEKLETLDGEEKAKVLNTPDALTDRYELIGIGPL 230
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 LPSFLQSKALPSALVTLREHIEALETESNPKILNTPFSALEHDALTVEKLMPIGPL 240

QY 231 IPSAFLDGGDPSETSYGGDLFEKSENNCNVEMWDTKPKSSVYVYVSGS-VLRFKQAMEE 289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 VSSS-----EGKT----DLFKSSDE--DYTKWDSKLSRSVIYISLGHADDLPEKHEA 289

QY 290 IGKGLLAGRPFLLWMIREQKNDGDEEEELSCIGELKMGKIVSWCSQLEVLHAHPALGC 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 290 LTHGVLATNREFLWIRE-KNPEKKKNRFLLELINGSR-GLVVGWCSQTAVLAHCAGCC 347

```

	QY	350	FVTHCGHNSAVESLSCGVPVVAVPQWFDQTNAKLIEDAWGTGVVRVMNEGGVDGSGSETE	409
	Dd	368	:          :          :          :          :          :	
	Dd	348	FVTHCGHNSLLESLESGVPVAFPOFADQCTTAKLVEDTTWRIGVKYKVGEGSDVDGEER	407
	QY	410	RCVEMVMDGCKSLVRENAIKWKTLAREAMGDGGSSLNKNLAFLHQ	456
	Dd	408	RCLKYMSGGEEAEEMRENAEKWKANAYDAAAGGPDNLNKGFDVE	454
			:    :    :    :    :    :    :    :    :    :    :	
	RESULT	6		
	Q9ZVY2		PRELIMINARY;	PRT; 519 AA.
	ID	Q9ZVY2		
	OC	Q9ZVY2:		
	DT	01-MAY-1999	(T=EMBLrel. 10, Created)	
	DT	01-MAY-1999	(T=EMBLrel. 10, Last sequence update)	
	DT	01-NOV-1999	(T=EMBLrel. 12, Last annotation update)	
	DE	T25N20.20		
	OS	Arabidopsis thaliana (Mouse-ear cress).		
	OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;		
	OC	core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;		
	OC	Arabidopsis.		
	RP	[1]		
	RR	SEQUENCE FROM N.A.		
	RA	SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,		
	RA	CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,		
	RA	VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,		
	RA	ECKER J.R.;		
	RT	*Genomic sequence for Arabidopsis thaliana BAC T25N20.*;		
	RL	Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.		
	RL	EMBL; AC005106; AAC80599.1;		
	DR	PROSITE; PS00375; UDPGT; 1.		
	SQ	SEQUENCE 519 AA; 58584 MW; C982007F CRC32;		

```

004930
ID      O04930          PRELIMINARY;           PRT;       474 AA.
AC      O04930;
DT      01-JUL-1997 (TReMBLrel. 04, Created)
DT      01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT      01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE      UDP-GLUCOSE:INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE
DE      (EC 2.4.1.121) (INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE).
GN      INAGLU.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      euclidyphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC      core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC      Arabidopsids.
RN      [1]
RN      SEQUENCE FROM N.A.
RP      RA THORNBERG R.W., GRAHAM R.A.;
RA      GILAD A., SCOLNIK P.A., BAR-ZVI D.;
RL      Plant Physiol. 113:1004-1004(1997).
RN      [2]
RN      SEQUENCE FROM N.A.
RA      THORNBERG R.W., GRAHAM R.A.;
RL      Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC      -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + INDOLE-3-ACETATE = UDP +
CC      INDOLE-3-ACETYL-BETA-1-D-GLUCOSE.
DR      ENBL; U81293; AAB58497.1; -.
DR      MENDEL; 16583; Arath;2543;16583.
DR      PROSITE; PS00375; UDPGT; 1.
DR      PFAM; PF00201; UDPGT; 1.
KW      Transferase; Glycosyltransferase.
SQ      SEQUENCE   474 AA;  53740 MW;  35521D70 CRC32;

Query Match              40.1%; Score 971; DB 10; Length 474;
Best Local Similarity    44.9%; Pred. No. 2.le-71;
Matches 216; Conservative 74; Mismatches 137; Indels 54; Gaps 17;

Qy  7 LLAFFPAAGHINPALQFAKRLK--AGTDVTFFTSVYAW-RMMANTASAAAGNPPGLDFV 63
Db  | |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  15 LVFVFPAAGHINSLEAKRLAGTISGARVTFAAISAYNNRMSTENV---PETLI 70

Qy  64 AFSDGYDGLKPC--GDGKR-----YMSEMKARGSEALRML-----LNHHVDVFPVVSH 111
Db  |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  71 TYSGDHDGDFKSSAYSDDKSQDATGFNFSEMRRRRGKETLTIELIEDNRKQNRPFCTCVVYTI 130

Qy  112 LFANAAEAVRESQVSALLWEPATVLICIVFYFNQYADEI----DAGSDEIQLPRLPPL 167
Db  | |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  131 LLTWAEARHFHPUSALLWQPVTVPFSIFTHYFNQYDAISEMANTPSSSIKLPSLL 190

Qy  168 EQRSFLPLLPTPERFL-MMKKELETLDGESEKAKVLNVFDALPDALTAI-DRYELI 225
Db  |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  191 TVRDIPSPVSSNVYAFLLPAFREQIDSLKEENPKILINTFOELEPEAMSSVPDNFKIV 250

Qy  226 GIGPLIPSAFLDGDPSTSGDLFEKSENNCVENWLDTPKPSVYVYFGSVLRFPKA 285
Db  ::::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  251 PVGPILL-----TLTDF-----SSRGEIEWLDTKADSSLVYSEFTLAVLSKK 294

Qy  286 QMEIEKGKLLACGRPFLMWIREQ--KNDDGEEEFELSCIGELKK---MGKIYSWCSQL 339
Db  |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  295 QLVELCALIQSRPFLWVITDKSYRNKEDQEKEE-DCISSEKSFDEIGNVYSWCQDF 353

Qy  340 EYLHAHALGCFTVTCGGWNNAVESLCGYPVVAPOWFQOTTNAKLIEDAQTGVRYV--RM 397
Db  |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  354 RVLNHRISGCVFTVTCGGWNSTLESVSGYPVVAFFQWMDMTNAKLEDCEWTGVRVMKK 413

Qy  398 NEGGS-GVDGSEIERCBWMVGDKSKLVRENATFKWTKLAREAMGEDGSSLKNLNAFLH 455
Db  --|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  414 EEGVVVVDSEIRRCIEVME--DKAEFRGNATRWDKDLAAEAVEREGSSFNHLKAFVD 471

Qy  456 Q 456
Db  ;
Db  472 E 472

```











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 12:39:50 ; Search time 9137.45 Seconds  
(without alignments)  
287.897 Million cell updates/sec

Title: US-09-147-955-3  
Perfect score: 1474  
Sequence: 1 accaaacccaaacaaattt.....acttaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 972840 seqs, 892348106 residues  
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0  
Maximum DB seq length: 1000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_om:\*
- 4: gb\_ov:\*
- 5: gb\_pat:\*
- 6: gb\_ph:\*
- 7: gb\_pl1:\*
- 8: gb\_pl2:\*
- 9: gb\_pr1:\*
- 10: gb\_pr2:\*
- 11: gb\_pr3:\*
- 12: gb\_ro:\*
- 13: gb\_sts:\*
- 14: gb\_sy:\*
- 15: gb\_un:\*
- 16: em\_fun:\*
- 17: em\_hum1:\*
- 18: em\_hum2:\*
- 19: em\_in:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_sy:\*
- 29: em\_un:\*
- 30: em\_vi:\*
- 31: gb\_htg1:\*
- 32: gb\_htg2:\*
- 33: gb\_in1:\*
- 34: gb\_in2:\*
- 35: em\_ba1:\*
- 36: em\_ba2:\*
- 37: em\_hum3:\*
- 38: em\_hum4:\*
- 39: gb\_pr4:\*
- 40: gb\_htg3:\*
- 41: gb\_htg4:\*
- 42: gb\_htg5:\*
- 43: gb\_htg6:\*

- 44: gb\_htg7:\*
- 45: em\_htg1:\*
- 46: em\_htg2:\*
- 47: em\_htg3:\*
- 48: em\_hum5:\*
- 49: gb\_pl3:\*
- 50: gb\_pr5:\*
- 51: gb\_htg8:\*
- 52: gb\_htg9:\*
- 53: gb\_htg10:\*
- 54: gb\_htg11:\*
- 55: gb\_htg12:\*
- 56: gb\_htg13:\*
- 57: gb\_htg14:\*
- 58: gb\_in3:\*
- 59: gb\_htg15:\*
- 60: gb\_htg16:\*
- 61: gb\_htg17:\*
- 62: em\_htg4:\*
- 63: em\_htg5:\*
- 64: em\_htg6:\*
- 65: em\_htg7:\*
- 66: em\_hum6:\*
- 67: gb\_htg18:\*
- 68: gb\_htg19:\*
- 69: gb\_htg20:\*
- 70: gb\_htg21:\*
- 71: gb\_htg22:\*
- 72: gb\_htg23:\*
- 73: gb\_htg24:\*
- 74: gb\_htg25:\*
- 75: gb\_htg26:\*
- 76: gb\_htg27:\*
- 77: gb\_htg28:\*
- 78: gb\_htg29:\*
- 79: gb\_htg30:\*
- 80: gb\_htg31:\*
- 81: gb\_vil:\*
- 82: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match			DB ID	Description
		Score	Length	DB		
1	1458	98.9	1458	7	AB013597	AB013597 Perilla f
2	1329.4	90.2	1507	7	AB013596	AB013596 Perilla f
3	680	46.1	2043	7	AB013598	AB013598 Verbena h
4	439.8	29.8	1864	7	AB027455	AB027455 Petunia x
C 5	284.4	19.3	198944	8	ATCHRIV38	AL161538 Arabidops
C 6	284.4	19.3	200576	7	ATFCA0	Z97335 Arabidops
7	262.4	17.8	1594	7	AB000623	AB000623 Nicotiana
C 8	226.8	15.4	84203	49	AC005106	AC005106 Genomic S
C 9	226.8	15.4	103223	8	AC007153	AC007153 Arabidops
C 10	186	12.6	195068	8	ATCHRIV42	AL161542 Arabidops
C 11	186	12.6	197419	8	ATCHRIV41	AL161541 Arabidops
C 12	186	12.6	205065	7	ATFCA4	Z97339 Arabidops
13	181.2	12.3	1589	8	ATU81293	U81293 Arabidops
14	142.4	9.7	1513	5	A62523	A62523 Sequence 28
15	142.4	9.7	1513	5	A72396	A72396 Sequence 28
16	138.8	9.4	99188	49	AC006533	AC006533 Arabidops
17	131.8	8.9	1588	5	A62526	A62526 Sequence 31
18	129.6	8.8	1459	5	A62525	A62525 Sequence 30
19	126.4	8.6	1969	5	A62529	A62529 Sequence 34
20	125.4	8.5	94786	49	AC002333	AC002333 Arabidops
21	124.6	8.5	1440	49	AF190634	AF190634 Nicotiana
22	121.6	8.2	1416	7	MZEIAGLU	L34847 zea mays IA
23	118	8.0	80818	7	AB019232	AB019232 Arabidops
24	116.8	7.9	566	5	A62509	A62509 Sequence 14











```

/ gene="AT4g14070"
complement(9562..9654,9750..9807,9883..9998,
10097..10176,10276..10435,10594..10731,11065..11145,
11247..11300,11383..11541,11653..11758,11882..11949,
12156..12305,12439..12554,12626..12834,13288..13360,
13958..14072,14674..14859,15291..15425))
/ note="strong similarity to AMP-binding protein - Brassica
napus; Contains Putative AMP-binding domain signature
AA255-266; Contains EST gb:A1999282.1"
/ codon_start=1
/ product="AMP-binding protein"
/ protein_id="CAB78449.1"
/ db_xref="GI:7268112"
/ translation="MASTSGSLSTLLSYGSPSPQFPDFGRFLISGHESVRIPSPRRF
REKEVPSPFLSSFSGDALRSSEKAVDPDIWSSAEKYGDRVALVDPYHDPDK
LIYKQLEQILDFAEGLRVLGKADKIALFADNSCRWLVSOGAVNVGRSSVVE
LQIYRHSSEVAIVDNPFFNRIAESFTSKASLRFLLILWGEKSLVTOGMQIPYIS
YAEINQGESRAKLSASNDTSYRNOFIDSDTAAIMYTSGETGNPKGYMLTHRLL
HQIKHLSKTVPAQAGKFLSMLPSWHAYERASEYFTTCGVEQMYTSIRYKDKLRY
QPNYIVSPLVETLYSGIQKISASSAGRKFLALTIKVSAMAYMEMKRIYEGMCLTK
EOKPPMYIAFVDMILARVIAALLWPLHMLAKLIYKKIHSSIGISKAGISGGSGLPI
HVDKFEALGYLQNGYGLTETSPVVCARTLSCNLVGSAGHPMHGTEFKIVDPETNNV
LPPSGGIIVRGVPQVMKGYIKNPSTCKVLNDSGFWNTGDTGWIAPHHSKGRSHCG
GTVLEGRADTIVLSTGENVEPLEIEAAMRSRVIEQIVIGQDRRLGAILIPNKE
EAGVDPERTSKETLSLVYQELRKWTSECSFQGPVLIVDDPFTIDNGLMTPTMKIRR
DMVAKYKKEIDOLYS"
complement(9655..9749)
/ number=1
complement(9750..9807)
/ number=2
complement(9808..9882)
/ number=2
complement(9883..9998)
/ number=3
complement(9999..10096)
/ number=3
complement(10097..10176)
/ number=4
complement(10177..10275)
/ number=4
complement(10276..10435)
/ number=5
complement(10436..10593)
/ number=5
complement(10594..10731)
/ number=6
complement(10732..11064)
/ number=6
complement(11065..11145)
/ number=7
complement(11146..11246)
/ number=7
complement(11247..11300)
/ number=8
complement(11301..11382)
/ number=8
complement(11383..11541)
/ number=9
complement(11542..11652)
/ number=9
complement(11653..11758)
/ number=10
complement(11759..11881)
/ number=10
complement(11882..11949)
/ number=11
complement(11950..12155)
/ number=11
complement(12156..12305)
/ number=12
complement(12306..12438)
/ number=12
complement(12439..12554)
/ number=13
complement(12555..12625)
/ number=13
complement(12626..12834)
/ number=14
complement(12835..13287)
/ number=14
complement(13288..13360)
/ number=15
complement(13361..13957)
/ number=15
complement(13958..14072)
/ number=16
complement(14073..14673)
/ number=16
complement(14674..14859)
/ number=17
complement(14860..15290)
/ number=17
complement(15291..15425)
/ number=18
complement(15426..16174)
/ number=18
/ gene="AT4g14080"
complement(join(16137..16174,16252..17604,17687..17732))
/ note="the cDNA A26668 is cognate to this gene; gene A6
for anther-specific protein - beta-1, 3-glucanase homolog;
Contains Glycosyl hydrolases family 17 signature
AA274-287; Prokaryotic membrane lipoprotein lipid
attachment site AA8-18; contains EST gb:AW004339.1,
Z18400"
/ codon_start=1
/ product="A6 anther-specific protein"
/ protein_id="CAB78450.1"
/ db_xref="GI:7268113"
/ translation="MSLLAFLEFTILVFSSCCSATRFQGHRYMQRKTMLDLASKIGI
NYGRGNLPSYQISNFIKIRAGHVKLVDAPESLTLLSOTNLVTVTPVNHQITA
LSSNQTIADEWFTNLPYVOTQIREVLGNELSYNSGVSNVLPAMRKIVNSIR
LHGHNKIVGTPLAMSLRSPFPNGTFFREETGVPMLPDLKFLNGTNSYFFLNVP
YFRWSRNPMTSLDFALFQGHSTYTDPTGLVYRNLLDQMLDSVLFAMTKLGYPHMRL
AISEGTWPNFGDIDETGANILNAATYRNRLIKKMSAPPIGTSPRGLPIPTFYVSLF
NENKSGSGTQRHWGILHPDGSPIYDVDTGTPLTGFNPLPKPTNNVYKGVQWCVF
VEGANETELETLMACAOQNTCAALAPGRECYEVPVSIYWHASVALNSYWAQPRQS
IQCFNGLAHEFTTNPENDRCRKPSTVL"
complement(16175..16251)
/ number=1
complement(16252..17604)
/ number=2
complement(17605..17686)
/ number=2
complement(17687..17732)
/ number=3
18189..18494
/ gene="AT4g14090"
/ number=1
18189..25761
/ gene="AT4g14090"
join(18189..18494,18694..18787,18938..19018,19571..19724,
19772..19833,21489..21524,22014..22792,22844..22948,
23060..23200,23280..23566,24375..24631,24686..25140,
25192..25311,25445..25614,25665..25761)
/ gene="AT4g14090"
/ note="similarity to proton-coupled peptide transporter
PEPT, Rattus norvegicus; Contains Prokaryotic membrane
lipoprotein lipid attachment site AA535-545; Prokaryotic
membrane lipoprotein lipid attachment site AA955-965;
contains EST gb:A1998193.1"

```

Query Match 19.3%; Score 284.4; DB 8; Length 198944;  
Best Local Similarity 56.2%; Pred. No. 3.3e-52;  
Matches 773; Conservative 0; Mismatches 511; Indels 92; Gaps 9;







misc_feature		/number=5 20514..20606 /note=*EM_ORG:CHSOPSBO Spinach plastid psbB operon with	
Query Match		19.3%; Score 284.4; DB 7; Length 200576;	
Best Local Similarity		56.2%; Pred. No. 3.3e-52;	
Matches 773; Conservative		0; Mismatches 511; Indels 92; Gaps 9;	
QY	48	tgtctgaacgcttcggtcgagccacataataatccccccctcccaatctcgccaagagac	107
DB	81653	TGCTTTACATATCCAGCGCAGGTCACATCAACCCGGCGCTTCAACTAGCAGCCGCC	81594
QY	108	tctataaagccgagcactgacgtcacgtttttcacgagcgttttatgatgcgagcgcatgg	167
DB	81593	TCATCCACACGCGTCAACCGTCACATATCTCCACCGCAGTCTCTGCTCACCAGCTATGG	81534
QY	168	ccaacacgacctcgccgctgcggtgcggaaccccaacccggcctcgactctcggtggttcg	227
DB	81533	GGAGCCACCTTCCA-----CAAAAGGTCTATCCTTCGCTTGGTTCACCG	81489
QY	228	acggtcacgacagcggtgaagccgcgagcgagcggaagcgctacatgtccgagatga	287
DB	81488	ATGGATTCCAGCAGCGTCTCAAGTCTATTGGAAGACCCAGAAAATCTACATGTCCGAACTCA	81429
QY	288	aagcccgcggtccgagcgcttaagaaacctctcttcaacacgacgacg-----	338
DB	81428	AACGATGTGTTCAACGCCCTCAGAGACATCATCAAGCCATCTTGACGCCACACCG	81369
QY	339	-----tcacttcgtctgaactccacacctcttgcattggcggtggaggtgg	386
DB	81368	AAACAGAGCTATCACCGGGTAATCTACTCTGTCTCGCTCGGTGGTTTCTACGGTAG	81309
QY	387	cgggtgttccacgctccgacgcccctctctggttcgagccgacccgctgctga	446
DB	81308	CGGTGAGTTTACCCTCCCACTACTCTCTCGGATTGAACGAGTACTGTACTAGACA	81249
QY	447	tatacactctacttcaacggtcagcagacgagatcagcgcggttccaatgaattc	506
DB	81248	TCCTACTACTACTTCAACACCTTTACAAACATCTCTTCGACGTT---GAACCGATT	81192
QY	507	agtcctcctgggttcacatccctggagcagcgagtgcttccgagcttctgctgcgcga	566
DB	81191	AATTACCGAACTGCCACTGATCACCACCGGTGACCTCCCGTCTCTTCTCAACCTTCA	81132
QY	567	cgcc---ggagagattccggttgatgaagagagagctggaacttttagacggtgaag	623
DB	81131	AGCATACCGTACGCTCTGTGACTCTTAAGAGAACATATCGAAGCTTCGAAACCGAAT	81072
QY	624	agaagcgaaagtattggtgaacacgcttttgatgcgttgagcccgatgcaactcagcgcta	683
DB	81071	CAAAACCTTAAGATTCTTTTAACACATCTCTGCTTTGGAACACGATGCTTTAACCTCTG	81012
QY	684	ttagatggtatgattggtgagatcggtgcccgttgatctccctcgcccttcttggacg	743
DB	81011	TTGAGAAACTCAAGATGATGCCAATCGGACCGTGTGTTCTCTCCGAGGGTAA-----	80955
QY	744	aagatccctccgaagacttaccgcgcgatcttttcaaaaaatcgaggagataaact	803
DB	80956	-----AACCGATCTTTTCAAAATCTTCCGACGAGGAT---T	80925
QY	804	gcgtgagtggttgaactgagccgcaaatcttcggtggtgtagtgcgtttggga---	860
DB	80924	ACAGAAATGTTAGATTCGAAGCTCGAGAGATCACTGATTTACATTTCTTTAGGCACAC	80865
QY	861	gcgttttgaggtttccaaagcgcaaatgggaagagattgggaagggctattagcctcg	920
DB	80864	ACGCCGATGATTTACAGAGAAACACATGGAAGCGTTACTACCGCGTGTAGCTACAA	80805
QY	921	gaagggcccttttatgatgatacagagaacagaagaatgacgacgcygaagaagaag	980
DB	80804	ACAGACCGTTTTTATGTCGTGAGGAGGAGAAAATCCAGAAG-----AGAAG	80758
QY	981	aagaagaagagtgagtgattgggggaattgaaaaaaatgggga-aaatagtgctggtg	1039

RESULT 7

AB000623	LOCUS	AB000623	1594 bp	mrna	PLN	05-FEB-1999
DEFINITION	Nicotiana tabacum mRNA for glucosyl transferase, complete cds.					
ACCESSION	AB000623					
VERSION	AB000623.1 GI:1805358					
KEYWORDS	JIGT; glucosyl transferase.					
SOURCE	Nicotiana tabacum					
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; Solanales; Solanaceae; Nicotiana.					
REFERENCE	1 (bases 1 to 1594)					
AUTHORS	Kojima, H.					
TITLE	Direct Submission					
JOURNAL	Submitted (23-JAN-1997) to the DDBJ/EMBL/GenBank databases. Hisae Kojima, school of Agriculture, Nagoya University, Laboratory of Biochemistry, Nagoya, Furo-cho, Chikusa 464-01, Japan (E-mail: hisae@nuagri.agr.nagoya-u.ac.jp, Tel: 052-789-4098, Fax: 052-789-4094)					
REFERENCE	2 (bases 1 to 1594)					
AUTHORS	Kojima, H., Hashizume, K., Imanishi, S. and Nakamura, K.					
TITLE	Jasmonate-induced potential glucosyltransferase from tobacco suspension cell					
JOURNAL	Unpublished (1997)					
FEATURES	Location/Qualifiers					
source	1..1594					
	/organism="Nicotiana tabacum"					
	/strain="BY-2"					
	/db_xref="taxon:4097"					
gene	/tissue_type="suspension culture"					
	52..1455					
	/gene="JIGT"					
	52..1455					
	/gene="JIGT"					
	/standard_name="jasmonate-induced potential glucosyltransferase"					
	/codon_start=1					
	/product="glucosyl transferase"					
	/protein_id="BAAL9155.1"					
	/db_xref="GI:1805359"					
	/translation="MENLKNCHVLIALFPQGHINPSLQFSKLLINLGKVTLLSSSL					
	SAFNRIKLPKIEGLTFAPFDGYNFGKSGFDDYHLNFAIKSHGSEFIANLISKA					
	KNQYPTFTVITILMDWAGSAVKLHPISTFLWQIATFYIYRTFNANFNKND					
	SODQIIEPLGPLSLSDPFPFVDVKSNDWAVESIKRQIEILNSENPRILVNTFD					







<http://compbio.ornl.gov/section/index.html>, GENSCAN (Chris Burge, & A.Salamov, Sanger Centre, <http://genomic.sanger.ac.uk/>), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/NetPlantGene.html>).

## FEATURES

Location/Qualifiers

```
1..103223
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F3F20"
1..33654
```

## misc\_feature

```
/note="overlap with bases 50550-84203 of 'TAMU' clone T25N20, gblAC005106. Features from bases 1 to 29,000 are annotated on overlapping clone T25N20, gblAC005106."
complement(29404..42021)
```

## gene

```
/gene="F3F20.1"
complement(join(29404..29652,29751..29936,30023..30073,30150..30950,31055..31284,31428..31486,31568..31651,31740..31858,31981..32100,32252..32411,32506..32695,32780..32848,32961..33086,33172..33281,33357..33508,33634..33700,33851..33948,34038..34161,34243..34403,34491..34575,35347..35441,35543..35680,35806..35892,35975..36024,36112..36220,36644..36981,37155..37272,37354..37473,37725..37818,37915..37991,38079..38204,38319..38445,38591..38682,38787..38879,38964..39080,39165..39245,39382..39505,39608..39729,39823..39923,40035..40201))
/gene="F3F20.1"
```

```
/note="Highly similar to putative callose synthase catalytic subunit; Highly similar to cotton putative
```

```
callose synthase catalytic subunit, gblAAD25952."
```

```
/codon_start=1
```

```
/protein_id="AAD30609.1"
```

```
/db_xref="gi:4836907"
```

```
/translation="MAQRREPPPPPPRRILRTOTVSGISGLAMLDSEVPPSSLSIAPILRVANEAGNPRVAYLCRFVAFKAHLRDLPTSSGGRVROFKTALQLRLENETTLAGROKSDARENQSFQHYKKYIQALLNAADKADRAOLTKAQTAQAVLFEVIAKVNQTEDEVEADRVNLIDDLIKTOIYYPNTPLDPSQNAIHRLEIQAQAAALNRQGLPWTAGHKHLLIDDLIQWQSMFGQKDNVLNREHLILLANVHIRQFPKPDQPKLLDRALTIYMKLLKWKYKGRKSLMLPTTQOQEQKRLLYMGLITLWGEAANDRPECUITYIHLYGKLAGSVPMTHGVKPAYGEQDEAFQKRVPIYQTSKEAKRSRGKSKHVNFRDNLNVEYFWSIRCFRLGWPMDADFCQTAEBELRSEIKNSGDWGMCKVFNFRSFIHFRSDRLWSEYILCLOAMIVIANNGSGLSALFOQDYLKVLVSPIITAILKLAQAVIDLALSKARHSMYLVKLRVMKVGAAVVMVAVTAYSNKASGFSQTIKNMFGHSHNSPSLFIIVAILIYLSPNMLSALLFLFFIRYLERSYKIMLMWMSQSYFPPSSMQIKPLVGTQDMIRHISVYSNHEEFPKAKNLGVIALWSPVILYFMDTOIYAIIVTLVGLNGAFRLGEIRTLNMLRSFRFQSIQAFNDCLVPQNSDDTKKRRPATESRKFDLPQSSKDEAARFQMWNNKIISSEFEEDLISDREMLLLVPYSDPDLDLIRWPPFLASKIPALDMDKDSKQBELKRLAVDSYMTCAVRECYASPKNLNLVVGREGQVINDIFSKIDHEIKETLITELNLSALDILQGFVRLEYLLENREEDKQDIVILVUNMLELVRTDIMEEVPNSANISVNDQSFILKRLKGGKKQIKRLHLLITKESAMDVPNSLEARRLTFPNSLMDPAPKIRNMLSFVLTPTFSEDLVSTIFGLEQNGEDGSILYLOKIFPDWNTFLERKVCQNEELRREDLBEELRLWASYRGQTLTKVRGMVYRKALEQAQDMADDELLKGYKALELTSEASKSGSLAQCALADMKFTFVYSCQYSIHRSQDQAKDILRLMTYPSIRVAYIDEVEQHKESYKTEKIIYYSALVKAAPQPKMDSSESQVTLQOLIYRIKLPAILGEGPEQNQNAIITRREGQITIDMNQDNMEEAFAKNNLQEELEKHGGRVRCPTILGLREHPTGVSSSLASMSNQNSFVTIGQRLVSLPKVRFYHPDIPDRLIFTRGFNSTLREGNVTHEYIQVKGDRDVLNQISMEAKIANGSEQTLSDRLYRLGHRFDRFMLSCYFTTIGFYETMLTVTVTVYVLYGRULVLVLEGLSQSRAFRNNKPLEAALASQSFVOIGFLMALPMMEIQLERGEFNALIEFVLMOLOLASVFTQLCTKHYGCTLFHGGAYRGTGRGVVFAKPAENRYFSRSHFVGIELMILLVYIQFGQSYRGVVTYILITVSMFMVTLWFAFLFNPSCFQWKIVDDMDWNKIYNRGGQVGPPEKSWESWKELEHLRHSVGRGTLLEIFLALREFIFQYGLVYHLSFTFKGNQSFWYGAQSFVIFLILLVGLRGSTGRFTNQFLFRIKGLVFLTFVAILITFLALPITTKDIFICMLAFWPTGMGLLIAQACKPLIQGLINSSVRLTARGYEIVMGHLLLFTFVAFLAWFPFVSEFQTMFLPNAFSEGLQISRIILGQQRDRSKNKE"
complement(44167..45812)
/gene="F3F20.2"
```

## gene

```
/complement(join(44167..44362,44438..44542,44760..45258,
```

## CDS

```
45331..45446,45524..45593,45680..45812))
/gene="F3F20.2"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD30611.1"
/db_xref="gi:4836909"
/translation="MESNGSGEVRVNVLYFLSRGSHVDHPLHLRVHLSRNGVFLRDKVKWLADARGDAMPDAFSGCKRRYKNGYQMDLDDLLTPIIDNRYVLKSGSILLSSPKEDYDNVEKKAWPTRNGGIDAEKLOKLKLTSEKIQKSPVCSORSTATTSTVTBESTINEGFFVLKKQDKPTVTSQGRDSTENGSDYSGRPSVSTTSSSYIKAKKSYSSVRASHVLRNLMKGGDLNDVAVLPLNKSNGAFPAWEDERYIQHQHNAKKSFEAGWSGIRKMETIEFKPKVAPSKPMAPLCSQCKGLFKPEKMHSHMLCKRGMKNSANNDLMTSNNTVKPQRCRNIPGNPLGHQRVLTTLKE"
complement(47118..50318)
/gene="F3F20.3"
complement(join(47118..48467,48761..49149,49186..49302,49523..49979,50064..50318))
/gene="F3F20.3"
/note="similar to Na/H antiporter proteins; N-terminal half of protein is similar to Na/H antiporter proteins"
/codon_start=1
/protein_id="AAD30610.1"
/db_xref="gi:4836908"
/translation="MSSGAPLVNPNYDIEESREKIVCYDQSLLFKREQKQWESGSLASSLFFITQIFVANLRYLYLPPFVAQILCGILFSPVSLGNTFRFIIAHVPRFTMTLETAFANLALVNFILGLGMDRLMRILTELKPVIIATITGLVALPVGAFLLYLPNGHPDKITISGCVFMSVALCTNPDRLARILADLKLSDMGRTAMCAIIVTDLCTWLLVYFGFASFSKSTWNKMPVITITAFVLLCIFVIRPGTAWTFAKTVKAGVLCGLDTCAGVHSITGAPLFGSLPHDHIIRNIEEKLHDSGLIMPLFIICGLRADIGFMLOFTDKPMVNVICSSFLVKIVTITSLFPHIMPRDAFALGMLNTKGLSVLVLNAGRDTKALDSPMYTHMTIALLVSLVVEPLAFAYKPKKLARYKHRTVQKIKGTELRLVACLVLPNVSGITNLLQVSNATKQSPLSVFAIHLVELTGRTTACLLI MNDECKPRANESDRVRAESDOIETFEAMVNDAMTQVTTAVSPYATMHEDICVLAEDKRYCFILIPYHKHLTDPGRMGEGNSHAELNQNVLSHAPCSVGLVDGRGMAYRSESPRGESMARVAMLFVGGPDDREALSYAMVMGQHVIKLTVVRFPVGRLEISSQKVAAYEREKQVDDCEIYEFNFKMTNDSVYIEKVNDGDTTATIREMEDNNSDYLVYGRGNSDPSVTAGLNDSSSPELGTIGDTLASSNTMHASVLVIOQYSATKRQAATAAAATTVMGAVAGVTCGNLESAGGDAKTRDAHEPPMKSMYDEDEDEHQQYGIHR"
51395..53210
/gene="F3F20.4"
join(51395..52184,52258..53210)
/gene="F3F20.4"
/note="similar to hexosaminidase; Similar to hexosaminidase, gi|1170249, gi|1076942 and others"
/codon_start=1
/protein_id="AAD30612.1"
/db_xref="gi:4836910"
/translation="MLTSLKSFHVLIPILFFITLSPLSFIALPINWPKRPLSWPQHKATALSNNFTILAPHEHQYLSASVTRYHNLIRSENYSPLISYPVKMLRYTLRNLVVTVDTSPLPHGVDESYSKLISPIGSPSAHLHLSAMGKLGLETFSQMTWGTSPDLCLPVGIXTODSPLGHRGVLLDTSRNYGVDDIMRTIKAMSANKLVNPHWHITDSQSPFLVLPSEPSLAAGSLGSDPMYTTPEDSKIVQYGEHGVRLPELDTGHTGSGEAYPEI VTCANMFPAKSWEEKLASEPGTQNLPLSPKTYEVVKNVQIDVYFPESFPFHGGDEVIPLGCKWTKDPAINSFLSGGTLSQLLEYINSTPIYVSNQNTVYVWEDVLLDAQIKADPSVLPKTLITQTNWNPENTKRIVAAAGRYIVSSSEFFYLDCCGGHGLNDSDIYDOKESGGSGCAPTKWQSYNYDIADGLLNEERKLVGLGEVALWESQADSTVLDLRLWPPASALASLNSGNERDGRVKRCGEAVDRLNRYMRVKRGIGAEPIQPLWCLNKPQMTNVHICALDQ"
53685..55199
/gene="F3F20.5"
53685..55199
/gene="F3F20.5"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD30626.1"
/db_xref="gi:4836924"
/translation="MSVVRVPRVLTPTSLLSQILKQKNPVLTALKLFEERKFPSPSYGH
```

## Query Match

15.4%; Score 226.8; DB 8; Length 103223;

Best Local Similarity 52.0%; Pred. No. 1.3e-39;

Matches 727; Conservative 0; Mismatches 602; Indels 69; Gaps 7;

QY 25 aaaaatggtccgcgcgctgctagcaacgtttccgcgcgaagccacataatcc 84

[illegible]

Db	18347	TGTTTTTTGACTCACTTTGGGTGGAGCTATCACTGGAGAGTTTGTTCTCGCGCTTC	18288
Qy	1138	gtggtggcgggtgccgcagtggttgatcacagcagcaaatgccaaagtattgagagatgc	1197
Db	18287	AGTGGTGGCTTTCCGATGTGTCGATCATGCCAGCAATGCCAAGCTTTTGAAGAAAT	18228
Qy	1198	gtgggggacaggggtgagagtgatgaatgaagggggtgggggttgatgagatgagat	1257
Db	18227	ATGGAAGACAGCTGTGAGGCTGAGAGAACTCGGAAGGTTTGTAGTAGAGAGAGAGAT	18168
Qy	1258	agaaaggtgtgagagatgtgatggatgggggtgacaagacccaactagtagagaaaa	1317
Db	18167	AATCGGCTTTTGGAAACGAGTGTATGGAGCG-----AATCGGTGGAGCTGAGGAAAA	18114
Qy	1318	tccatcaaatgaagactttggcagacaaagccatgggatgagatggatcttcactcaa	1377
Db	18113	CCGAGAAATCGAAGCGTTTACGACTACAGCGGTAGAGAGGAGGATCTTCGCACAA	18054
Qy	1378	caatctcaacgcctttct	1395
Db	18053	GAATGTGGAAGCTTTGT	18036
RESULT	10		
LOCUS	ATCHRV42/c		
DEFINITION	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 42.	PLN	16-MAR-2000
ACCESSION	AL161542		
VERSION	AL161542.2	GI:7268303	
KEYWORDS	thale cress.		
SOURCE	Arabidopsis thaliana		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi		
REFERENCE	1 (bases 1 to 195068)		
AUTHORS	EU Arabidopsis sequencing, project.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: John Bevan, Centre for Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk		
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a> this fragment has an overlap with ATCHRV41 at the 5' end and an overlap with ATCHRV43 at the 3' end.		
FEATURES	Location/Qualifiers		
source	1..195068		
	/organism="Arabidopsis thaliana"		
	/variety="Columbia"		
	/db_xref="taxon:3702"		
exon	/chromosome="4"		
	9917..10066		
	/gene="AT4g15560"		
	/number=1		
CDS	join(9917..10066,10350..10475,10559..10874,10981..11252,11340..11613,11691..11998,12081..12362,12441..12725,12813..12953)		
	/gene="AT4g15560"		
	/note="GT/AC sites ignored to preserve continuity; light inducible gene required for chloroplast development; Contains Prokaryotic membrane lipoprotein lipid attachment site AA452-462; Prokaryotic membrane lipoprotein lipid attachment site AA591-601; Transketolase signatures AA102-121; Transketolase signatures AA505-521; contains EST gb:AW04332.1, AI997953.1, W43562, N96015, W43388"		
	/codon_start=1		
	/product="DEF (CLAI) protein"		
	/protein_id="CAB78598.1"		

/db\_xref="GI:7268304"  
/translation="MASSAFAPPSYIIITKGGUSTDSCKSTSLSSRSLSVTDLPSPCLK  
PNNSHNRKAVCASLAKEGGEYYSNRPTPLDITNYPIMKMLNSVKELKOLSELR  
SDVIFNVSKTGHLSSGVLVELTVALHYIFNTPODKILMDVGHOSYPKHITGRGRK  
MPTMRGTGLSGFTRGSEHDCFGTGHSSSTISAGLGMVAGRDLDKNNNNVAVIGD  
GAMTAGQAYEAMNAGYLDSDMIVILNDNKQVSLPTATLDGSPVPVGLSLSRLQS  
NPALRELREAVAKMTQIGGPMHQIAAKVDEYARGMISGTSLSFEELGLYYIGPVDG  
HNIDVAILLKVKSTRTGPVLIHVTEKGRGYPAERADDKYHGVMKFDPATGRQF  
KTKNTQSTVTVFEALVAEAEVDKDVVAIHAAMGGTGLNLFQREPTRCFDVGIAE  
QHAYTFAGLAGEGLKPCAIYSSFEQRAYDQVHDVDLQKLPVRFAMDRAGLVGADG  
PHCGAFDVFMACLPIMVIMAPSDAEDLFNVAATAVADIDDRPSCFRYPVPRNGIGVAL  
PPNGVPTIEGKRLKEGREGVALLGYGSVOSCGAAVLEERGLNVTADARFC  
PLDRLAIRSLAKSHVELTIEGSGGSHVYQFLALDGLDGLDKLRPMVLPDRIY  
DHGAPADOLAEAGLMPSHIAATNALNIGAPREAL"  
9917..12953  
/gene="AT4g15560"  
10067..10349  
/gene="AT4g15560"  
/number=1  
10350..10475  
/gene="AT4g15560"  
/number=2  
10476..10558  
/gene="AT4g15560"  
/number=2  
10559..10874  
/gene="AT4g15560"  
/number=3  
10875..10980  
/gene="AT4g15560"  
/number=3  
10981..11252  
/gene="AT4g15560"  
/number=4  
11253..11339  
/gene="AT4g15560"  
/number=4  
11340..11613  
/gene="AT4g15560"  
/number=5  
11614..11690  
/gene="AT4g15560"  
/number=5  
11691..11998  
/gene="AT4g15560"  
/number=6  
11999..12080  
/gene="AT4g15560"  
/number=6  
12081..12362  
/gene="AT4g15560"  
/number=7  
12363..12440  
/gene="AT4g15560"  
/number=7  
12441..12725  
/gene="AT4g15560"  
/number=8  
12726..12812  
/gene="AT4g15560"  
/number=8  
12813..12953  
/gene="AT4g15560"  
/number=9  
14413..14436  
20338..24557  
/gene="AT4g15570"  
20338..20373  
/gene="AT4g15570"  
/number=1  
Join(<20338..20373,20767..20852,20962..21012,21256..21367,  
21467..21572,21653..21750,21989..22117,22182..22274,  
22625..22678,22957..23015,23122..23206,23286..23765,

24183..24269,24366..24557)  
/gene="AT4g15570"  
/note="similarity to SEN1 protein, yeast; Contains  
ATP/GTP-binding site motif A (p-loop) AA13-20; contains  
EST gb:N65040"  
/codon\_start=1  
/product="SEN1 like protein"  
/protein\_id="CAB78599.1"  
/db\_xref="GI:7268305"  
/translation="VGLSRKSPVLIOGPCTGTOTILSILGAIHMTAPRVQSKGTD  
HEVKGIOMTIOEKYHNHGRASPWILGVNPRDAIMPEDGDDGFPTSGNELKPEVYNA  
SKYRLRVLCAPSNSALDEIVRLSSGLRDENAQTYTKIVRIGLKAHHSVAVSL  
DLHVAQKRGSALDKPKQGTGTDIDISRTAILEEAAIVMGFNFRINVFATLSFGSAL  
LAKSNRGFDVLIIDEAAQAVPATLIPLATRCQVFLGDKPQLPATVISTVAQDSGY  
CTSMERLQKAGYPVKMLKTOYRMHPELIPISQPKIVVSLIPMTIASYETLQTFI  
LTENQIRSPFSQKFEGALEDDSDIEAQTTROWHKYRCGPGCFDDIHEGKESQHPGA  
TGSRVNLDIEVEFLLIYHRLVTMYPKLSSQLAIIISPYNQVTKFKDKEMFGTEA  
EKVDINTVDGVGVGSAATLKSDDLKMLNKLIESAQNRNLKFKVSKPLNNFFSENLLET  
MKLTEDMEIPDAPLYEDESPLVAPYGGDDDFGDDADQDDVANAGED"  
20374..20766  
/gene="AT4g15570"  
/number=1  
20767..20852  
/gene="AT4g15570"  
/number=2  
20853..20961  
/gene="AT4g15570"  
/number=2  
20962..21012  
/gene="AT4g15570"  
/number=3  
21013..21255  
/gene="AT4g15570"  
/number=3  
21256..21367  
/gene="AT4g15570"  
/number=4  
21368..21466  
/gene="AT4g15570"  
/number=4  
21467..21572  
/gene="AT4g15570"  
/number=5  
21573..21652  
/gene="AT4g15570"  
/number=5  
21653..21750  
/gene="AT4g15570"  
/number=6  
21751..21988  
/gene="AT4g15570"  
/number=6  
21989..22117  
/gene="AT4g15570"  
/number=7  
22118..22181  
/gene="AT4g15570"  
/number=7  
22182..22274  
/gene="AT4g15570"  
/number=8  
22275..22624  
/gene="AT4g15570"  
/number=8  
22625..22678  
/gene="AT4g15570"  
/number=9  
22679..22956  
/gene="AT4g15570"  
/number=9  
22957..23015  
/gene="AT4g15570"  
/number=10



```

intron      23016..23121
            /gene="AT4g15570"
            /number=10
exon        23122..23206
            /gene="AT4g15570"
            /number=11

Query Match      12.6%; Score 186; DB 8; Length 195068;
Best Local Similarity 59.2%; Pred. No. 1.2e-30;
Matches 362; Conservative 0; Mismatches 235; Indels 15; Gaps 2;

Qy 799 taactgctgagtggttgactgaagcgaacatcttcggtgtgtatgtgtcgtttg 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4203 TGAATACATAGAGTGGTGGTACTAAAGCGGATTCGTCTGCTTTATGTTCTGTCGG 4144

Qy 859 gagcggtttgagttccaaaggcacaaatggaagatgaggaaagggctattagcctg 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4143 GACCGTTGCCGTGTGTGACGAAGAACAGCTTGTGGAGCTTTGTAAAGCGTTGATACAAAG 4084

Qy 919 cggaggcccttttggatgatacagaaacag-----aagaatgacgacgcgga 969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4083 TCGGAGACCATCTCTGTGGGTGATTACGGATAAGTCGTACAGAAATTAAGAGATGAGCA 4024

Qy 970 agaagaagaagaagaagagtggtgctgattgggggaattgaaaaaaatggggaaaat 1029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4023 AGAAGAAGGAAGATTCATAGTACTTTTCAGAGAAGAGCTCGATGAGATGAGAAATGGT 3964

Qy 1030 agtctggtgctcgagtggtgaggttctgagcaccctcgcttggtgagatgcttcgtagc 1089
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3963 GGTTCATGTTGTTGATCAGTTTAGGGTTTGAATCATAGATCGATAGTGTGTTTCGTGAC 3904

Qy 1090 gcattggtgggaactcgctgtggaagcgttgattgctgaggtatcccggtggtggtcgtt 1149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3903 GCATTGGGGTGGAACTCTACGCTGGAGAGCTTGTTTCAGAGCTTCGGTGGTGGCGTT 3844

Qy 1150 gccagtggttttcagacagcgaatgcgaagctgattgagatgctggtggggagcagg 1209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3843 TCCGAATGGAATGATCAGATGATGAACGCCGAAGCTTTTAGAAGATTTTGGAAAAACAGG 3784

Qy 1210 ggtgagagt-----gagaatgaatgaaggggtggtggtggtgagatgagatagaag 1263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3783 TGTAAAGTGTATGAGAGAGGAAGAGAGAGAGTGTGTGTGTGATAGTAGGAGAT 3724

Qy 1264 gtgtggaagtgtgattgagtggtgacaaagcacaactagtggagaaaaatgccat 1323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3723 ACGCGGTGCATTAGGAAGTATTGGAAGACAAAGCGGAGAGGTTTGAAGAAATGCCAC 3664

Qy 1324 caaatggaagactttgcccagacaagccatgggtagatggtgattcttcaactcaacaatct 1383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3663 GAGGTGAAGGATTTAGCGCGGAGGCTGTGAGAGAGGAGGCTCTTCCTTTAATCATCT 3604

Qy 1384 caacgcctttct 1395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3603 CAAAGCTTTTGT 3592

RESULT 11
ATCHRIV41/c
LOCUS      ATCHRIV41 197419 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 41.
ACCESSION AL161541
VERSION    AL161541.2 GI:7268270
KEYWORDS   thale cress.
SOURCE     Arabidopsis thaliana
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsi.
REFERENCE  1 (bases 1 to 197419)
AUTHORS   EU Arabidopsis sequencing, project.
TITLE     Direct Submission
JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer

Biochemie, An Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mps.biochem.mpg.de,mayer@mps.biochem.mpg.de,Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV40 at the 5' end and an
overlap with ATCHRIV42 at the 3' end.
Location/Qualifiers
1..197419
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
/number=1
complement(join(21638..22378,22677..23498))
/note="similarity to hypothetical 66.0 kd protein
F47D12.7; contains EST gb:AA712915"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB78566.1"
/db_xref="GI:7268271"
/translation="MNQTRRRPDSLSRLDLSPEMLFPKSPSTSSIRNIFLL
LIFCIIFIYSTGNFREQISLSPVSVKDYRFPYFPGGLRAISARVRYKET
LWTSPESTRVAVFLDRLGLESDLTPVIVSKDYRFPYFPGGLRAISARVRYKET
FVGKGFPAISLAKVLAKLVDSCLMRYSHMVGSDSRIFSCVAELGVLTHEPGPHQID
VRNIFGLCAHPLSPLYSLHLDVDPFPKRNRTSEVAHLGAASFDGRILQOVS
RCDLNTVTVVWVGVAQVTEGNKLLDLTLQKTFSTFRGSGVQSNFMFSTRETP
RDCPGRLVFFLSDSGDTGEGTWSNINIRHVGCHRAEAVERLERIRVLSKLERNV
EQVGFINTDRFSVKPSILPGTFFFGFLKIKQMGLPFRQCCDISSPYNKSMVINIR
QCMPELIANT"
complement(22379..22676)
/number=1
complement(22677..23498)
/number=2
28265..29226
/gene="AT4g15250"
complement(join(28265..28651,28738..29226))
/note="similarity to CONSTANS protein, Arabidopsis
thaliana"
/codon_start=1
/product="CONSTANS like protein"
/protein_id="CAB78567.1"
/db_xref="GI:7268272"
/translation="MEARCDFCGTEKALIIYCKSDSAKLCLNCDVNVHSANPLSORHTR
SLICERSLQPTAVHGMNENSLCCGQWNTASNTGLGHRQLQSLNIPSDCPSDFGR
IWSSTLEPSTLSVSPFSDTLQLQELDDNGSSSTVQTQNLKDYSSFFMESNLPKV
IEECSGLDCEGINLNDAPLNFNASNDIIGCSLSDNTKCYEYEDSFEKENNIGLPSL
LLPTLSGNVYNNMSLSMSNLTGSRNATDYDCGISPGFLIGDSPWESNVESFNPCLR
DEAKRYKOKKSKRM"
complement(28265..28651)
/number=1
complement(28652..28737)
/number=1
complement(28738..29226)
/number=2
31839..31950
/gene="AT4g15260"
/number=1
31839..33264
/gene="AT4g15260"
join(31839..31950,32018..33264)
/gene="AT4g15260"
/note="strong similarity to UTP-glucose
glucosyltransferase; Contains UDP-glucosyltransferases
signature AA319-362; contains EST gb:T20820, T46737,
AI993247.1, AI994260.1, AI996826.1, AI997955.1, T76043,
```



Db 196202 TGTAAAGAGTGTATGGAGAAAGAAAGAAAGAGTTGTGGTGGCTAGTGCAGGAGAT 196143  
 QY 1264 gtgtgtggagatggtgatgatgggggtgacaagaccacaactagttagagaaaaatgccat 1323  
 Db 196142 ACGGCGGTGCATTTGAGGAAGTTATGAAGAACAAGCGGAGGAGTTTAGAGAAATGCCAC 196083  
 QY 1324 caaatggaagacitttggccagacaagccatgggataggatggaatcttctactcaacaatct 1383  
 Db 196082 GAGGTGGAAGGATTTAGCGCGGAGGCTGTGAGAGAAGGAGGCTCTTCTTAAATCATCT 196023  
 QY 1384 caacgcctttct 1395  
 Db 196022 CAAAGCTTTTGT 196011

RESULT	12
ATFCM4/c	
LOCUS	ATFCM4 205065 bp DNA PLN
DEFINITION	Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment No. 4.
ACCESSION	297339
VERSION	297339.2 GI:5281015
KEYWORDS	.
SOURCE	thale cress..

ORGANISM

Arabidopsis thaliana  
Eukaryota: Viridiplantae  
euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons: core  
eudicots: Rosidae: eurosids II: Brassicales; Brassicaceae;  
Arabidopsids.  
1 (bases 1 to 205065)

REFERENCE  
AUTHORS

Bevan, M., Stekema, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N.,  
Kreits, M., Kavanagh, T., Entian, K.D., Rieger, M., James, R.,  
Puidumenech, P., Hatzipoulos, P., Obermaier, B., Duesterhoft, A.,  
Jones, J., Palme, K., Ansorge, W., Delsen, M., Bancroft, I.,  
Mewes, H.W., Schueller, C. and Chalhatzis, N.

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 205065)
AUTHORS	EU Arabidopsis sequencing project.
TITLE	Direct Submission
JOURNAL	Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk
COMMENT	On Jun 30, 1999 this sequence version replaced gi:2244901. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a> this fragment has an overlap with ATFFCA3 at the 5' end and an overlap with ATFFCA5 at the 3' end.

```

FEATURES             Location/Qualifiers
     source            1..205065
                        /organism="Arabidopsis thaliana"
                        /variety="Columbia"
                        /db_xref="taxon:3702"
                        /chromosome="4"
     source            1..14029
                        /organism="Arabidopsis thaliana"
                        /db_xref="taxon:3702"
     gene              /clone="cosmid YUP5B11.IB1G15.YUP5B11.IB1G19"
                        4257..7108
                        /gene="dl3766w"
     exon              4257..4662
                        /gene="dl3766w"
                        /number=1
     CDS               join(<4257..4662,6078..7108)
                        /gene="dl3766w"
                        /note="strong similarity to fatty acid hydroperoxide lyase
                        - Capsicum annum; contains EST gb:949688"
                        /codon_start=1
                        /product="hydroperoxide lyase (HPOL) like protein"

```

[illegible]

```

10065..10511
/gene="dl3770c"
join(12306..12597,13233..13495,13554..13835,13944..14000,
15045..15290,15396..15802,16095..16280,16380..16453,
17015..17346)
/gene="dl3775w"
/notes="similar to hypothetical protein C02F5.7,
Caenorhabditis elegans; contains EST
gb:1217311.937779;315469"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB10325.1"
/db_xref="GI:2244904"
/db_xref="SPTREMBL:O23399"
/translation="MDKPYGYASVSMGIDRSAGKDIDLEMVGVEATLYPGLSYGPNQ
LRGFTKRYKYGILSAQLLTLLISAVVNLNPPVNDLLTSGPGLILFLCIVPILLSFE
FVSTEDLVLSHRSRTAIFYNSILTLFRFAHVIDYWEFNFWIWPILHIHQHPVNLII
LIAFLVSVLSFTVGVSCAMTEGCIPLNPSYCHKRONLPETWNSKACKSCVFTNLLI
FLLRPVIGRIVLOALILTLSVGSILTATFWAAKKGKDFSLGPLTFLSLIILVVTLS
FTQMPPLGPTVAVYVGGFSALKRRGRDSSPSKRRKLLDKTHSGAENVESSTLS
DTGLALANGFRIENLSLIWCPNVSSVGLCSLAOKCTSLKSLDLOGCVVGQGLAIV
GFECKLELNLRFEGLTDVGVIDLVCSSKSLKSGVAAAKITDLSLBAVGSCK
LLEVLDSLEYIHDKGLIAVAGCHRLNKLQCVSDVAFVAAGELCTSLERLALY
SFQHFYDKMRAIGKSGKKLQDLTSDCYFVCKGLEAIAHGCKELERVEIHGCHNIG
TRGIEAIGKSCPLKELALLYCQRIQNSALQIEIGKYLKAGTFDHRFQVIGDWPLAEL
GGCPMLKDLVLSHCHITDGNLNLVQKCKLLETHMVYCPGITSAGVATVYVSSCPH
IKKVLTEKKVTERITRRAGSYISLYCMLD"
12306..12597
/gene="dl3775w"
/number=1
12306..17346
/gene="dl3775w"
12598..13232
/gene="dl3775w"
/number=1
13233..13495
/gene="dl3775w"
/number=2
13496..13553
/gene="dl3775w"
/number=2
13554..13835
/gene="dl3775w"
/number=3
13836..13943
/gene="dl3775w"
/number=3
13944..14000
/gene="dl3775w"
/number=4
14001..15044
/gene="dl3775w"
/number=4
14700..147605
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="cosmid YUP5B11.IB112/YUP5B11.IB1A8"
15045..15290
/gene="dl3775w"
/number=5
15291..15395
/gene="dl3775w"
/number=5
15396..15802
/gene="dl3775w"
/number=6
15803..16094
/gene="dl3775w"
/number=6
16095..16280
/gene="dl3775w"
/number=7

```

```

Query Match 12.6%; Score 186; DB 7; Length 205065;
Best Local Similarity 59.2%; Pred. No. 1.2e-30;
Matches 362; Conservative 0; Mismatches 235; Indels 15; Gaps 2;

QY 799 taactcggtgagtggtgaactcgaagccgaatacttcgggtggtgtatgtcgtttg 858
Db 47149 TGAATACATAGAGTGGTTGGATACTAAAGCGGATTCTGCTGTCTTATCTTCGTCGG 47090

QY 859 gacggtttgaggtttcccaaaagccacaaatggaagagattgggaaagggctatagcctg 918
Db 47089 GACGCTTGGCGTGTGTAGCAAGAAACAGCTTGTGGAGCTTTGTAAAGCGTGTATACAAAG 47030

QY 919 cggaaggccctttttatggtatgatacaggaacag-----aagaatgacgacggcga 969
Db 47029 TCGGAGACCATTCTTGTGGGTGATTACGGATAAGTCTTACAGAAATAAAGAAATGAGCA 46970

QY 970 agaagaagaagaagaagaagtggtgagttgcatcattgggggaattgaaaaaaatggggaaaaat 1029
Db 46969 AGAAGAAGAAAGAGATTGCATAAGTAGCTTTTCAGAGAAGAGCTCGATGAGATAGGAATGGT 46910

QY 1030 agtgcgtggtgctcgagttgaggttcttcggcgaccctgcttgggatgtttcgtgac 1089
Db 46909 GGTTCATGGTGTGATCAGTTCAGGGTTTGAATCATAGATCGATAGGTGTTTCGGTGAC 46850

QY 1090 gcatttggttggaactcgctgtggagagcttgagttcggggattccggtcggcggtg 1149
Db 46849 GCATTGCGGGTGGNACTCTACGCTGGAGAGCTTGGTTTCAGGAGTTCGGGTGGCGCTT 46790

QY 1150 gccgcagtggtttgatcagacgacgaatgcgaagctgattgaggtgcgttgggggacag 1209
Db 46789 TCCGCAATGGAATGATCAGATGATGAACGCGAAGCTTTTAGAAGATTGTTGGAAACAGG 46730

QY 1210 ggtgagagt-----gagaatgaatgaaggggtgggttgatggatgagtagataaag 1263
Db 46729 TGTAAAGAGTGTATGAGAAAGAAAGAAAGAGGAGTTGTTGGTGGTGGATGAGGAGAT 46670

QY 1264 gttgtggagatggtgatggtggtggaagacaaactagtgagagaaaatgccat 1323
Db 46669 ACCGCGTGCATTGAGGAAGTTATGGAAGACAGAGCGGAGGAGTTTACAGGAATGCCAC 46610

QY 1324 caaatggaagactttggccagacaaagccatgggatggatggatcttcactcaaatct 1383
Db 46609 GAGGTGGAAGATTATAGCGCGGAGGCTGTGAGAGAAGGAGGCTCTTCCCTTAATCATCT 46550

QY 1384 caacgctttct 1395
Db 46549 CAAAGCTTTTGT 46538

RESULT 13
ATU81293
LOCUS
DEFINITION
  Arabidopsis thaliana UDP-glucose:indole-3-acetate
  beta-D-glucosyltransferase (laglu) mRNA, complete cds.
ACCESSION
  U81293
VERSION
  U81293.1 GI:2149126
KEYWORDS
  thale cress.
SOURCE
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
  eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
  Arabidopsi.
REFERENCE
  1 (bases 1 to 1589)
  Graham, R.A. and Thornburg, R.W.
  DNA Sequence of UDP Glucose:Indole-3-acetate
  Beta-D-Glucosyltransferase from Arabidopsis thaliana (Accession No.
  U81293) (PGR97-044)
  Plant Physiol. 113, 1004 (1997)
REFERENCE
  2 (bases 1 to 1589)
  Thornburg, R.W. and Graham, R.A.
  Direct Submission
  Submitted (06-DEC-1996) Biochemistry and Biophysics, Iowa State

```

```
University, Ames, IA 50011, USA
FEATURES
  source            1..1589
                    /organism="Arabidopsis thaliana"
                    /variety="Columbia"
                    /db_xref="taxon:3702"
  gene              26..1589
                    /gene="iaгу"
  CDS               26..1450
                    /gene="iaгу"
                    /EC_number="2.4.1.121"
                    /function="transfers the glucose moiety from UDP-glucose
                    to indole-3-acetic acid to form ester linked IAA-Glc"
                    /note="IAA-glucose synthase"
                    /codon_start=1
                    /product="UDP-glucose:indole-3-acetate
                    beta-D-glucosyltransferase"
                    /protein_id="AA058497.1"
                    /db_xref="GI:2149127"
                    /translation="MANNNNSPTGPHLFVTPAOGHINPSLELAKRLAGTISGARV
                    TFAASAYNRMFSTENVPETLIFATYSDDHDDPKSSAYSDDKSDODANGNEMSEMR
                    RKCELTLEIEDNRKONRPTCVVYTLITWVAELAREHLPALLWVQPVFVSIF
                    YHFNGETDAISEMANTPSSINKLPSLLTVDIPSVSSNNYAFLLPFAFREQIDS
                    LKEEINPKILINTFOELEPAMSVDPNFKIVPVPLLTLRDTSRGEYIEWLDTKA
                    DSVLYVSGTGLAVLKKQLVELCKALIQRRPFLMWITDKSYRNKDEDEKEDCIS
                    SSKSPDEIGMYSWCDQFRVLNHRISGCVPTHCWNSTLESVLVSGVPVVAFFQWNDQ
                    MTNAKLLDCWKVTVRVMKKKEEGVVVVDSEIRICIEBEMEDKAEFRGNATRWKD
                    LAEAVREGSSFNHLKAFDEHM"
  polyA_signal      1568..1575
                    /gene="iaгу"
  polyA_site        1588..1589
                    /gene="iaгу"
                    /note="24 A nucleotides"
  BASE COUNT       433 a 342 c 375 g 439 t
  ORIGIN
Query Match      12.3%; Score 181.2; DB 8; Length 1589;
Best Local Similarity 58.7%; Pred. No. 1.2e-29;
Matches 35; Conservative 0; Mismatches 238; Indels 15; Gaps 2;

Qy 799 taactgcgtgagtggtgaactcgaagccgaaatcttcggtggtgatgctgctgttg 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 823 TGAATACATAGAGTGTGGGATCTAAGCGGATTCGTCGTGCTTTATGTTTCGTTGG 882
Qy 859 gacgcttggagtttccaaagccaaatgaagagattgggaagggctattacgctg 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 GACGCTTGCCGTGTGAGCAAGAACAGCTTGTGGAGCTTTGTAAAGCGTTGATACAAAG 942
Qy 919 cggagggcccttttatggatgatacgaagacag-----aagaatgacgagcgga 969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 TCGAGAGACCATCTTGTGGGTGATTACGGATAAGTCGTACAGAAATAAAGAGATGAGCA 1002
Qy 970 agaagaagaagaagaagagtggtgattgctgggaattgaaaaatggggaataat 1029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1003 AGAAGAAGAACAGATGTCATAAGTAGTTCAGAGAAGAGCTTCGATGATGATGAGTAATG 1062
Qy 1030 agtctgctgctgcagattgaggttcttcgagcaccctcgttggagatttcgtgac 1089
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1063 GGTTCATGTTGTTGATCAAAATTAGGGTTTGAATCATAGATCGATAGTGTGTTTCGTGAC 1122
Qy 1090 gcattgtgggtggaactcgtgtgagagccttgattgcgggattccggtggtgagcgtt 1149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1123 GCATTGGGGTGGAACTCTACGCTGGAGAGCTTGGTTTCAGAGATTCCGGTGGTGGCGTT 1182
Qy 1150 gccgagtggtttgatcagacgacgaatgcgaagctgattgagatgcgtgggggacagg 1209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1183 TCCCAATGGAATGATCAGATGACGAACGCGAAGCTTTTAGAAGATTGTTGGAACACAGG 1242
Qy 1210 ggtgagagt-----gagaaatgaatgaaggggttggggttggatggtgagatgagaag 1263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1243 TGTAAGATGTGAGAGAAGGAAGAGAGAGTGTGTGTGGTGGATAGTGGAGAGAT 1302

Qy 1264 gtgtgtgagatggtgatgggtgggtgagacaagcaaacactagtgagagaaatgccat 1323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1303 ACGGGGTGCATTTAGGAAGATTATGGAAGAACAGCGGAGGAGTTTATGAGGAATGCCAC 1362
Qy 1324 caaatgaagactttggccagacaagccatgggatagatgattgattcactcaacaatct 1383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1363 GAGGTGGAAGCATTTAGCGGGGAGGCTGTGAGAGAGAGGAGGCTCTCTCTTTAATCACT 1422
Qy 1384 caacgcctttct 1395
    ||| ||| ||| |||
Db 1423 CAAAGCTTTTGT 1434
    ||| ||| ||| |||

RESULT 14
LOCUS      A62523      1513 bp      DNA      PAT      09-MAR-1998
DEFINITION Sequence 28 from Patent WO9716559.
ACCESSION  A62523
VERSION     A62523.1 GI:3716430
KEYWORDS   .
SOURCE      unidentified.
ORGANISM   unidentified.
REFERENCE   1 (bases 1 to 1513)
AUTHORS    Van,A.K., Marillia,E., Peferoen,M., Grootwassink,J.W., Reed,D.W.,
            Hemmingsen,S.M., Kolenovsky,A.D., Underhill,E.W. and
            Macpherson,J.M.
TITLE      Plants with reduced glucosinolate content
JOURNAL    Patent: WO 9716559-A 28 09-MAY-1997;
            PLANT GENETIC SYSTEMS NV (BE)
FEATURES   Location/Qualifiers
            source          1..1513
            /organism="unidentified"
            /db_xref="taxon:32644"
            69..1469
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAA03659.1"
            /db_xref="GI:3716431"
            /translation="MAETTTTTATNSKGHVVLVPYVQGHLPVMOFAKRLVSKGVK
            VTIATTTTASSISTPVSVEPISDGHDHFDLIGVGVSIDAYSSEFKLNGSETLTVI
            SKFKSTDPIDSLVDYDFLPWGLVARSNSIAAAFTNNLTVCVSRKRVSGGFPLP
            ADASAPLYGLPALSYDELPSGVSRHSSHAHGRVLLNPNRHHEDADLVFNVSFE
            GLTQCEGVESEAMRATLIGPMIPSAVLDGRIKDDKGYGSSLMKPLSESCMEWLDTK
            LSKSVFVSFGSGILFEKQLAEVAKALQESNPNFLVIREAHIAKLPEGFVEATKDR
            ALLVSWCNQLEVLAHVSIIGFLTHCGWNSTLEGSLGVPMVGVQWSDQMDAKFVEE
            VWRVGYRAKEEGGVVKSDEVVRLGRVMEGESSVEIRESSKKWKDLAVKAMSEGG
            SDRSINEFVESLKKH"
            SDRSINEFVESLKKH"
BASE COUNT 368 a 356 c 399 g 390 t
ORIGIN
Query Match      9.7%; Score 142.4; DB 5; Length 1513;
Best Local Similarity 47.2%; Pred. No. 3.7e-21;
Matches 634; Conservative 0; Mismatches 666; Indels 42; Gaps 5;

Qy 39 gccgcgtgctctagcaacgcttccgcgcgaagccacataaataccccccctcccaattcg 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112 GCCACGCTCGTGGTCTTACCTTACCAGTCCAGGCCACCTCAACCCAAATGGTCCAGTTCC 171
Qy 99 ccaagagactcctaaagccgcgactgacgtcaagctttttccacgagctttatgatggc 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 CTAAACGCTAGTCTCCAAAGCGGTCAAGTCAACAATGCCACCACCACCTTACACCGCT 231
Qy 159 gccgcattggcacaacagcctcccgctgcgcggaaccccccccggtcgtacttcgtgg 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 CCTCCATCTCCACTCCCTCCGCTCTCCGTCGAACCAATCTCCGACGGCCACGACTTCATCC 291
Qy 219 cgttctccagcgtctagcaacgagcgggtgaagccgcgcgcgcgcgcgcgcgcgcgcgc 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 TCATAGGCGTCCCGGGCTCAGCATCGACGATAGTCCGGAATCTCTCAAGCTCAACGCT 351
Qy 279 ccgagatgaagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 338
```





**THIS PAGE BLANK (USTO)**









Qy	531	agcagcgagctcttcgacagctttctgctgcctgcgacccgagagattccgggtgatga	590
Db	540	CCACGCGTGATTTACCGCTCTTCCCTTCTCCATCCGCATGAGAGATTCCGTTCACTGA	599
Qy	591	tgaagagaaagcttggaaacttttagacggtgaagagaaaggcgaagatttggfgaacacgt	650
Db	600	TGAAGGAGAAATTTGGAACATTTAGAAAGGTGAAGAAAAACCTTAAGGCTCTTGGTGAACAGCT	659
Qy	651	ttatgcgttggagcccgatgcactcaagcgtattgatagattatgattatcgggatcg	710
Db	660	TTGATGCGTTTGGAGCTGATGCGCTCAAGCCATTTGATAAGTACGATGATTGCAATCG	719
Qy	711	ggcgcgttgattccctccgctcttctggacggcgaagatccctccgaacgcttccacggcg	770
Db	720	GGCGGTTGATTCCTCCGATCTTTGGACGGTAAAGATCCCTCGGACAGGCTTTTCGGCG	779
Qy	771	gcgatcttttcgaaaaa----tcggaggagaaataactcgtcgtagtggttgaactcgaagc	827
Db	780	GAGATTTGTTTCGAGAAAGGCTCCAATGACGAGATTTGCCTCGAATGTTTGAGCACGAATC	839
Qy	828	cgaactctcggfggfgtatgctgcttttggagagcgttttgaggtttcccaaaagcaca	887
Db	840	CTCGATCTTCGGTGGTTACGTTTCGTCGGAAGCTTCGTTAAACGACGAAGTCGCAAA	899
Qy	888	tggaaagatttgggaaaggctattagcctcgggaagccctttttatggatgatacgag	947
Db	900	TGGAAGAGATAGCAACAGGGCTGTAGATTGTGGGAGCCGCTTTTGTGGGTGGTAAAG	958
Qy	948	aacgaagaatgcacgcgcgcgaagaagaagaagaagaagagtgtagttgcattgggg	1007
Db	959	-----GTAACGAAGGAGAAAGAGGTATTGATAAGTTGATTCATGCGAGG	998
Qy	1008	aattgaaaaaaatggggaaaaatagtcgcgtgctgcagttcggaagttctcgcgcacc	1067
Db	999	AGTTGAAACAGATGGGGAAATTTGATCTCTGTTGTTCTCAATTGGAAGTCTCGACGCATC	1058
Qy	1068	ctgcgttggatatttcgtgacgcatttggtggaactcgcgtggaagacttgagtt	1127
Db	1059	CCTCGTTGGGATGTTTCGTGACACATCGCGGTGGAAATTCACATCTAGAGAGTATATCTT	1118
Qy	1128	gcgggattccggtgggtggcggcagtcggttgatcagacacacgaatcgagctga	1187
Db	1119	TCGGGTTCCGATGGTGGCTTTTCCGCAGTGGTTTCGATCAAGGACGAAATCCGAAGCTGA	1178
Qy	1188	ttgaggtgcggtggggacaggggtgagagtgagaaatgaatgaagggggt---gggggtg	1244
Db	1179	TGGAGGATGTGTGGAGACCGGGTGTGAGAGTCAGAGCTAAATGACGAGGGTAGCGCTTG	1238
Qy	1245	atggatgagataganaaagtgctgtggagatggtgatggatgggggtgacaagaccacaac	1304
Db	1239	ATGGTGATGAATTTAGGAGATGATTTTGGAGAGGGTTATCGATGGGGGAGAAAAGTAGTA	1298
Qy	1305	tagtgagagaaatgcacataatggaacactttggccagacaaagccatggatagatg	1364
Db	1299	AACCTAGAGAGAGTGTGGCAAGTGGAGAGATTTGGCAAGAAAGACTATAGGAGGAAGTG	1358
Qy	1365	gattctcactcaacaactctcaacgcctttcttcgtccaagttgcacacttttaactctgctc	1424
Db	1359	GATCTTCAGTTAACAACCTCAAGGTCCTTCTTGATGAGGTTGTAGGTATCTTAAGACGTA	1418
Qy	1425	aaacag 1430	
Db	1419	AATGAG 1424	

## RESULT 4

RESULT  
X02829

X02829  
ID X02829 standard; cDNA; 1671 BP.

AC X02829;

DT 14-MAY-1999 (first

DE W09905287 Seg ID 4.  
KW Plant; flavonoid 5-transglycosylation activity; SMG; variety; colour; ds.





```
FT conflict acid change"
FT /tag= y
FT /note= "G is C in clone pGL3-22; there is no amino
FT acid change"
FT 678. .679
FT /tag= z
FT /note= "bases 678-679 are AT in clone pGL3-22,
FT causing an amino acid change from Arg-204
FT to Ile-204"
FT 685
FT /tag= aa
FT /note= "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 714
FT /tag= ab
FT /note= "A is G in clones pGL3-22 and pGL4-2,
FT causing an amino acid change from Ser-216
FT to Gly-216"
FT 759
FT /tag= ac
FT /note= "G is A in clone pGL3-22, causing an amino
FT acid change from Ala-232 to Thr-232"
FT 761
FT /tag= ad
FT /note= "G is C in clone pGL3-22, causing an amino
FT acid change from Ala-232 to Thr-232"
FT 766
FT /tag= ae
FT /note= "G is A in clones pGL3-22 and pGL4-2,
FT causing an amino acid change from
FT Arg-234 to Lys-234"
FT 794
FT /tag= af
FT /note= "A is C in clone pGL3-22; there is no amino
FT acid change"
FT 809
FT /tag= ag
FT /note= "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 814
FT /tag= ah
FT /note= "G is C in clone pGL3-22, causing an amino
FT acid change from Gly-249 to Ala-249"
FT 848
FT /tag= ai
FT /note= "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 901
FT /tag= aj
FT /note= "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 936
FT /tag= ak
FT /note= "G is A in clone pGL3-22, causing an amino
FT acid change from Gly-290 to Arg-290"
FT 959
FT /tag= al
FT /note= "C is A in clone pGL3-22; there is no amino
FT acid change"
FT 968
FT /tag= am
FT /note= "G is A in clone pGL3-22; there is no amino
FT acid change"
FT 973
FT /tag= an
FT /note= "A is C in clone pGL3-22, causing an amino
FT acid change from Lys-302 to Thr-302"
FT 1013
FT /tag= ao
FT /note= "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 1117
```

```
FT /tag= ap
FT /note= "G is T in clones pGL2-7 and pGL2-25,
FT causing an amino acid change from
FT Val-350 to Glu-350"
FT 1121
FT /tag= aq
FT /note= "G is T in clone pGL2-7; there is no amino
FT acid change"
FT 1130
FT /tag= ar
FT /note= "C is T in clone pGL2-7; there is no amino
FT acid change"
FT 1161
FT /tag= as
FT /note= "T is C in clone pGL2-7; there is no amino
FT acid change"
FT 1169
FT /tag= at
FT /note= "A is G in clone pGL2-7; there is no amino
FT acid change"
FT 1193
FT /tag= au
FT /note= "T is G in clone pGL2-7; there is no amino
FT acid change"
FT 1202
FT conflict

Query Match 9.7%; Score 142.4; DB 1; Length 1513;
Best Local Similarity 47.2%; Pred. No. 1.2e-28;
Matches 634; Conservative 0; Mismatches 666; Indels 42; Gaps 5;

Qy 39 gccggtgctagcaagcttccgcgagccacataatccgcctccattcg 98
Db 112 GCCAGTCTGCTTACCTTACCAGTCCAGGCCACTCAACCAATGTCGATTGCG 171
Qy 99 ccaagactactaaagccgactgactgactgttttcaacgagcgttatgcatg 158
Db 172 CTAACGGCTAGTCTCCAAAGCGTCAAAAGTCAACATGCCACCACCACTACCGCT 231
Qy 159 gccgcatggcaacacagctccgcgctccggaacccacccgggctcgactcg 218
Db 232 CCTCCATCTCCACTCCCTCCGCTCTCCGTCGAACCAATTCGCAGCGCCAGCTTCATCC 291
Qy 219 cgttcgcgaggtacgaacgaggggtgaagccgagcgagcgagggagcgactc 278
Db 292 TCATAGGCGTCCCGCGCTCAGCATCGAGCATATCCGAATCTTCAGCTCAAGGCT 351
Qy 279 ccgagatgaagccgagcggtccgagggccttaagaaacctctctcaaacagcag 338
Db 352 CCGAAACCTTAACCGAGTAATCTCAAAATTCACAGATTCACCCATCGA----- 406
Qy 339 tcaatttcgtctactccacacctctttgcatggcgagggcgaggttgctgccc 398
Db 407 ----TTCATTAGTCTACGACTCTTTCTCCGTCGGGACTCGAAGTCGAGATCTAACT 462
Qy 399 acgtcccgagccctctctcgggtcgagcgccgacccgctgctgt---gcataccact 455
Db 463 CCATCTCAGCTGCTGCTTCTTCACCAACACCTCACCGCTTGTCTGTTCACCAAT 522
Qy 456 tctactcaacggttaacgacagagatcgacgagcggttccaatgaattcagctccctc 515
Db 523 TCGTCTCCGGTGAGTTTCTCTCCCGCTGATCCCGCTTCGCGCGGTATCTCGTCGGTG 582
Qy 516 ggcttccatccctggagcagcagctcttcgacagcttctgctgctcgagccggaga 575
Db 583 GCTTACCGGCTTTGAGCTACGACGAGCTTCTCTCTCGTGGAGCTCAGCTCGTGAGCC 642
Qy 576 gattccggtgatgatgaagagaagctggaacctttagagcggtgaagaggaag 635
Db 643 ACGGGAAACACGGGAGAGTTCTCTGAACACAGTTCCCGTACCAACAGAGATGCTGATTGGC 702
Qy 636 tattggtgaacacgtttgatgcgttggagcccgatgcactcagcgttatgtaggtatg 695
Db 703 TGTTCGTCAACAGCTTCGAAGGTTAGAGACACAGAGTTGTCAAGTTGGAGAAATCAGAGG 762
```

QY 696 agtggatgc-----ggatggccggttattccctccgccttcttggagcggcgaagatc 749  
 Db 763 CGATGAGCGGACGTTGATCGGACCTATGATACCATCTGTTATCTCGACGCGCGAATCA 822  
 QY 750 cctccgaacgcttctacgcgccgcatcttttcgaaaaatcgaggagagaataactcgtgg 809  
 Db 823 AAGACGATAAAGGCTACGCTCGAGCTGAT---GAAGCCGCTCTCGGAGGAGTGTATGG 879  
 QY 810 agtgggtgaactcgaacgcaaatcttcggtggtgtagtgcgttggagcgttttga 869  
 Db 880 AGTGGTTAGACACTAAGCTGACCAAGCTCGGTGTTTTCGTTGTTGCTTCTTTGGGA 939  
 QY 870 ggtttccaaagccacaaatggaagagatcgggaagggcgtattagcctcgggaagccct 929  
 Db 940 TCCTCTTCGGAAGCAACTCGCTGAGGTGGCAAGGCGTTACAGAAATCAACTTTAACT 999  
 QY 930 tttatggatgatacagaaacgaagaatgacgacgcgcaagaagaagaagaagaag 989  
 Db 1000 TCTTGTGGGTGATCAAGAAGCTCATATAGCGAAGTTACAGAAAGGTTTGTGGAAGCTA 1059  
 QY 990 agttagtgcattgggaatgaaataatgggaaataatgtagtgcgtggtgcgcagt 1049  
 Db 1060 CC-----AAAGCAGACGCTGCTGTTCTTCTTGTGTTAACACAGC 1098  
 QY 1050 tggaggttctggcgccacctgcgttgggtggtggtggtggtggtggtggtggtggtggt 1109  
 Db 1099 TTGAGGTTTGTAGCTATGATGATAGTGTGCTTTTGTACTACTCGGTTTGGAACTCGA 1158  
 QY 1110 ctgtggagagcttgatgctgggagatcccggtggtggtggtggtggtggtggtggtggt 1169  
 Db 1159 CGTTGGAAGGATGAGTTTGGGAGTTCCGATGTTGCTGCTGCTGCTGCTGCTGCTGCTG 1218  
 QY 1170 cgaagaaatcgaagctgattgagatgagatgagatgagatgagatgagatgagatgag 1229  
 Db 1219 TGAATGATCTAAGTTTGTGGAGAGGTTTGGAGAGTTGGGTATAGAGCAAGGAGGAAG 1278  
 QY 1230 aagggggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1289  
 Db 1279 CTGGGGGAGGAGTTTGAAGACGATGAGTGTGAGGTTTGAAGAGGATGATGGAAG 1338  
 QY 1290 gtgacaagaccaaactagtggagagaaatgccatcaaatggaagacatttggccagacaag 1349  
 Db 1339 GAGAGAGTAGTGTGGAGATTAGAGAGAGTTCTAAGAAGTGGAAAGATTGGCTGTCAAGG 1398  
 QY 1350 caatggagtaggtagtatcttc 1371  
 Db 1399 CGATGAGTGAAGGAGGAAGCTC 1420

RESULT 7  
 T10106  
 ID T10106 standard; cdna to mRNA; 1731 BP.  
 AC T10106;  
 DT 27-OCT-1996 (first entry)  
 DE Maize UDP-glucose indol-3-yl:acetyl:glucosyl transferase sequence.  
 KW UDP-glucose:indol-3-yl:acetyl:glucosyltransferase; enzyme;  
 KW transgenic plant; EC-2.4.1.121; crop improvement; corn; cereal;  
 KW grass; IAGlu-transferase; ds.  
 OS Zea mays.  
 FH Key  
 FT cds  
 FT misc\_binding  
 FT Location/Qualifiers  
 FT 57..1472  
 FT /\*tag= a  
 FT /\*tag= b  
 FT /\*note= "start of UDP binding region"  
 FT  
 PN WO9600291-A1.  
 PD 04-JAN-1996.  
 PR 20-JUN-1995; U07820.  
 PF 24-JUN-1994; US-265427.  
 PA (RESE) RESEARCH CORP TECHNOLOGIES INC.  
 PI Banduraki RS, Szczeglowski K, Szerzen JB;  
 DR WPI; 96-068875/07.

DR P-PSDB; R99999.  
 PT UDP-glucose indol-3-yl:acetyl:glucosyl transferase DNA and protein  
 PT - used in sense or anti-sense orientation for the genetic control of  
 PT plant rAA levels and plant growth  
 PS Claim 2; Page 21-23; 4lpp; English.  
 CC This nucleotide sequence encodes a UDP-glucose-indol-3-yl-  
 CC ylacetylglucosyltransferase (IAGlu-transferase) from maize (Zea mays).  
 CC Over-expression of this sequence in transgenic plants will result  
 CC in at least 1 of the following effects: inhibition of apical  
 CC dominance; inhibition of stem elongation; inhibition of cell  
 CC enlargement and increased numbers of stems per plant as compared  
 CC with a wild-type plant. The ability to control the expression of  
 CC this enzyme allows the control of free indoleacetic acid levels in  
 CC plants, thereby effecting plant growth rates. Antisense constructs  
 CC may be used to inhibit synthesis of this enzyme.  
 CC Sequence 1731 BP; 298 A; 528 C; 601 G; 304 T;  
 SQ  
 Query Match 8.2%; Score 121.6; DB 1; Length 1731;  
 Best Local Similarity 52.8%; Pred. No. 4.5e-23;  
 Matches 321; Conservative 0; Mismatches 269; Indels 18; Gaps 2;  
 QY 793 ggagaataactcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 852  
 Db 848 GGAGGATGCGTGCACCAAGTGGCTAGACACACCAAGCCGACCGCTCGCTGCGCTACGTC 907  
 QY 853 gtttggagcgttttgggttttccaaagcacaataatggaagagattgggaagggtatt 912  
 Db 908 CTTTCGCGACGCTCGCTGCTGCTGCGCAACGCCAGAGAGGAGGAGCTCGCGCGCGCTCCT 967  
 QY 913 agcttcggaagggccctttttatggtggtggtggtggtggtggtggtggtggtggtggt 1012  
 Db 968 CGCCGCGCGCAAGCCCTTCTGTTGGGTGGTGAGGGC-----CAGCCACGA 1012  
 QY 973 agaagaagaagaagaagattggtggtggtggtggtggtggtggtggtggtggtggtggt 1032  
 Db 1013 GCACAGGTCCCGCGCTATCTCTGCGCGAGCGACGGCGCGCGCGATGTCGT 1072  
 QY 1033 gtcgtggtcgtcagttggaggttctgcgcacccctgctggtggtggtggtggtggtggt 1092  
 Db 1073 GCCTGCTGCTGCGCGACGCTGACGCTGCTGCGCGCGCGCGCGCGCTGCTGCTGCTG 1132  
 QY 1093 ttgtggtggaactcgtcgtggtggtggtggtggtggtggtggtggtggtggtggtggt 1152  
 Db 1133 CTGCGGTTGGAATCTCCAGCTGAGGCGCTCAGCTTCGCGCTGCTGCTGCTGCTGCTG 1192  
 QY 1153 gaagtgtttgatcagacgacgaatgcgaagctgattgaggtgctggtggtggtggtggt 1212  
 Db 1193 GCTGTGGAGCGGACGCT 1252  
 QY 1213 gagagtgaagaatgaagaagg---gggtgggtggtggtggtggtggtggtggtggtggt 1269  
 Db 1253 GCT 1312  
 QY 1270 ggagatggtgattggtgggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1329  
 Db 1313 GCGCGCGCTCATGCGCGCGCGCGCGCGCGCGCGCTGCTGTCGACGCAAGCGCGCGGGA 1372  
 QY 1330 gaagatttggccagacaagcgaatgggtaggtggtggtggtggtggtggtggtggtggt 1389  
 Db 1373 GAGGAGACAGGCTCGCT 1432  
 QY 1390 cttttttt 1397  
 Db 1433 GTTCGTGC 1440  
 RESULT 8  
 T66173  
 ID T66173 standard; cdna to mRNA; 566 BP.  
 AC T66173;  
 DT 15-JUL-1997 (first entry)  
 DE UDP-glucose:thiohydroximate S-glucosyltransferase clone pGL2-7.



KW Glucosinolate; UDP-glucose:thiohydroximate S-glucosyltransferase;  
 KW S-GT; transgenic plant; rapeseed oil; oilseed rape; canola; ds.  
 OS Brassica napus cv. Westar.  
 PN EP-771878-Al.

PD 07-MAY-1997.  
 PF 31-OCT-1995; 402425.  
 PR 31-OCT-1995; EP-402425.  
 PA (CAN) ) NAT RES COUNCIL CANADA.  
 PA (PLB2 ) PLANT GENETIC SYSTEMS NV.  
 PI Grootwassink JWD, Hemmingsen SM, Kolenovsky AD, Peferoen M;  
 PI Reed DW, Underhill EW, Van Audenhove K;  
 DR WPI; 97-247418/23.

PT Plants genetically transformed to interfere with  
 PT UDP-glucose:thiohydroximate S-glucosyltransferase gene expression  
 PT - useful for production of rapeseed oil with reduced glucosinolate  
 PT content

PS Example 2; Page 18-19; 35pp; English.  
 CC Brassica napus UDP-glucose:thiohydroximate S-glucosyltransferase  
 CC (S-GT) partial cDNA clones pGL2-7 (T66173) and pGL2-25 (T66174)  
 CC were obt'd. by PCR-RACE (see also T66167-72) of B. napus cDNA.  
 CC Analysis of the sequences revealed an open reading frame of about  
 CC 470 bp, a 104 3' untranslated region and polyA+ tail for both  
 CC clones. The amino acid sequence of the encoded protein revealed  
 CC part of S-GT peptide 2 (W09827), as expected because this sequence  
 CC was used for PCR-cloning, and the complete S-GT peptides 5, 6 (with  
 CC one amino acid difference) and 7 (W09830-32). A full-length cDNA  
 CC clone (T66166) for B. napus S-GT (W09825) was subsequently obt'd.  
 SQ Sequence 566 BP; 160 A; 54 C; 193 G; 159 T;

Query Match 7.9%; Score 116.8; DB 1; Length 566;  
 Best Local Similarity 56.5%; Pred. No. 5.4e-22;  
 Matches 217; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy 1013 aaaaaatggggaaaaatagtcgtggtgctgcgcagttggaggttctcggcgacacctgcg 1072  
 Db 47 AAAGACAGAGCGTTGCTTGTGTTCTTGGTGTAAACAGCTTGAGGTTTACGTCATGGATCT 106

Qy 1073 ttgggatatttcgtagcaccattggtggtggaactcgtgctgcagttggaggttctcggcgacacctgcg 1132  
 Db 107 ATAGGTTGTTTTTGTGACCTACCTGCGGTTGGAACTCGAGCTTGAGGTTTACGTCATGGATCT 166

Qy 1133 attccggtgggtggcgagcagttggtggtggaactcgtgctgcagttggaggttctcggcgacacctgcg 1192  
 Db 167 GTTCCGATGGTGGGTGTCGCCAGCTGGAGTGATCAGATGAATGATGCTTAAGTTTGTGGAG 226

Qy 1193 gatcggtgggggacaggggtgagagtgagaaatgaatgaaggggtggtggttgatgagatgt 1252  
 Db 227 GAGGTTTGGAGAGTTGGGTATAGAGCGAAAGATGAAGCTGGGGGAGGAGTTGTGAAGAGC 286

Qy 1253 gagatagaaagtgtgtgagatggtgatggatgggttgacagacaaactaagtgaga 1312  
 Db 287 GATGAGGTGGAGAGTTGGGTATAGAGCGAAAGATGAAGCTGGGGGAGGAGTTGTGAAGAGC 346

Qy 1313 gaaaatgccatcaaatggaagactttggccagacacatgggtaggtaggtaggtcttca 1372  
 Db 347 GAAGGTTCTAAGAAATGGAAGATTGGCTGTGAAGGCGATGAGTGAAGGAGGAGCTCT 406

Qy 1373 ctcaacaatctcaacgcctttctt 1396

Db 407 GATCGGAGCATTAATGAGTTGTT 430

RESULT 9  
 T66174

ID T66174 standard; cDNA to mRNA; 568 BP.

AC T66174;

DT 15-JUL-1997 (first entry)

DE UDP-glucose:thiohydroximate S-glucosyltransferase clone pGL2-25.

KW Glucosinolate; UDP-glucose:thiohydroximate S-glucosyltransferase;

KW S-GT; transgenic plant; rapeseed oil; oilseed rape; canola; ds.

OS Brassica napus cv. Westar.

PN EP-771878-Al.

PD 07-MAY-1997.  
 PF 31-OCT-1995; 402425.  
 PR 31-OCT-1995; EP-402425.  
 PA (CAN) ) NAT RES COUNCIL CANADA.  
 PA (PLB2 ) PLANT GENETIC SYSTEMS NV.  
 PI Grootwassink JWD, Hemmingsen SM, Kolenovsky AD, Peferoen M;  
 PI Reed DW, Underhill EW, Van Audenhove K;  
 DR WPI; 97-247418/23.  
 PT Plants genetically transformed to interfere with  
 PT UDP-glucose:thiohydroximate S-glucosyltransferase gene expression  
 PT - useful for production of rapeseed oil with reduced glucosinolate  
 PT content

PS Example 2; Page 19; 35pp; English.  
 CC Brassica napus UDP-glucose:thiohydroximate S-glucosyltransferase  
 CC (S-GT) partial cDNA clones pGL2-7 (T66173) and pGL2-25 (T66174)  
 CC were obt'd. by PCR-RACE (see also T66167-72) of B. napus cDNA.  
 CC Analysis of the sequences revealed an open reading frame of about  
 CC 470 bp, a 104 3' untranslated region and polyA+ tail for both  
 CC clones. The amino acid sequence of the encoded protein revealed  
 CC part of S-GT peptide 2 (W09827), as expected because this sequence  
 CC was used for PCR-cloning, and the complete S-GT peptides 5, 6 (with  
 CC one amino acid difference) and 7 (W09830-32). A full-length cDNA  
 CC clone (T66166) for B. napus S-GT (W09825) was subsequently obt'd.  
 SQ Sequence 568 BP; 157 A; 55 C; 197 G; 159 T;

Query Match 7.4%; Score 109.4; DB 1; Length 568;  
 Best Local Similarity 55.4%; Pred. No. 5.1e-20;  
 Matches 212; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

Qy 1013 aaaaaatggggaaaaatagtcgtggtgctgcgcagttggaggttctcggcgacacctgcg 1072  
 Db 47 AAAGACAGAGCGTTGCTTGTGTTCTTGGTGTAAACAGCTTGAGGTTTACGTCATGATCG 106

Qy 1073 ttgggatatttcgtagcaccattggtggtggaactcgtgctgcagttggaggttctcggcgacacctgcg 1132  
 Db 107 ATAGGTTGTTTTTGTGACCTACCTGCGGTTGGAACTCGAGCTTGAGGTTTACGTCATGATCG 166

Qy 1133 attccggtgggtggcgagcagttggtggtggaactcgtgctgcagttggaggttctcggcgacacctgcg 1192  
 Db 167 GTTCCGATGGTGGGTGTCGCCAGCTGGAGTGATCAGATGAATGATGCTTAAGTTTGTGGAG 226

Qy 1193 gatcggtgggggacaggggtgagagtgagaaatgaatgaaggggtggtggttgatgagatgt 1252  
 Db 227 GAGGTTTGGAGAGTTGGGTATAGAGCGAAAGAGGAGGAGGAGGAGTTGTGAAGAGC 286

Qy 1253 gagatagaaagtgtgtgagatggtgatggatgggttgacagacaaactaagttaga 1312  
 Db 287 GATGAGGTGGAGAGTTGGGTATAGAGCGAGGAGTTGTGAAGGAGGAGTAGTGTGGAGATTAGA 346

Qy 1313 gaaaatgccatcaaatggaagactttggccagacacatgggtaggtaggtaggtcttca 1372  
 Db 347 GAGAGTTCTAAGAAATGGAAGATTGGCTGTGAAGGCGATGAGTGAAGGAGGAGCTCT 406

Qy 1373 ctcaacaatctcaacgcctttctt 1395

Db 407 GATCGGAGCATTAATGAGTTTCT 429

RESULT 10  
 T68693

ID T68693 standard; cDNA; 432 BP.

AC T68693;

DT 05-JAN-1998 (first entry)

DE Strawberry UDP-glucuronosyl transferase cDNA.

KW Strawberry; UDP-glucuronosyl transferase; ERT1b; fruit; ripening;

KW storage; gene regulation; ss.

OS Fragaria ananassa Duch. cv. Brighton.

PN W09721816-Al.

PD 19-JUN-1997.

PF 12-DEC-1996; G03076.

PR 13-DEC-1995; GB-025459.

PA (ZENE) ZENECA LTD.

PI Manning K;  
 DR WPI: 97-332787/30.  
 PT Vectors encoding new enzymes for regulating ripening of fruit - used  
 PT particularly to improve storage properties, processing  
 PT characteristics etc, especially in strawberries  
 PS Claim 1: Page 23-24; 43pp; English.  
 CC This cDNA sequence comprises a UDP-glucuronosyl transferase (ERT  
 CC 1b) cDNA clone isolated from a ripe strawberry fruit cDNA library  
 CC by differential screening. 9 Ripening-enhanced clones (see T68687-  
 CC 95) clones were isolated from the library, and are deposited at  
 CC NCIMB 40693. Any of these 9 sequences can be used as a regulatory  
 CC sequence in claimed vectors for genetic transformation of plant  
 CC cells to regulate fruit ripening. The vectors also comprise a  
 CC promoter and transcription termination sequence. Also claimed are  
 CC plants (especially strawberry) and propagating materials containing  
 CC these vectors. Stable integration of the regulatory sequences, or  
 CC their complements or genomic equivalents, into a plant is used to  
 CC modulate fruit ripening by overexpression or downregulation of an  
 CC endogenous plant gene. Slowing ripening of fruit improves  
 CC resistance to damage during harvesting, packaging and transport.  
 CC It also extends shelf life, improves storage properties, processing  
 CC characteristics, flavour and aroma, and modifies colour and  
 CC increases resistance to post-harvest fungi.  
 CC Sequence 432 BP; 101 A; 84 C; 136 G; 94 T;  
 SQ

Query Match 5.1%; Score 75.8; DB 1; Length 432;  
 Best Local Similarity 60.4%; Pred. No. 4.2e-11;  
 Matches 125; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 1015 aaaaatggggaaaatagtcgtggtgcgcagttggaggttctggtgcacccctgcgtt 1074  
 DB 105 AGACAGAGCAAAAGTTGTCAATGGAGTCCACAAGAGAGATTTTGGACATCTTCGAC 164

QY 1075 gggatgttcgtgcagcagcattggtggaactcgtggtgagagcttgagttgcggat 1134  
 DB 165 GGCTTGCTTTGTACTCATTTGCGGTGGAACTCAACCATGGAGTCACTCACCTCAGGAAT 224

QY 1135 tccggtggtggcgggtgccagtggttgcagcagcagcagcagcagcagcagcagcagc 1194  
 DB 225 GCCCGTGGTGGCATTCACACATGGGGTGCACCAAGTGCAGCGCCCAAGTATTTGTCGA 284

QY 1195 tgcgtgggggacagggtgagagtgag 1221  
 DB 285 CGAGTTTAAGTGGGAGTAAGAATGTG 311

RESULT 11  
 VI7054  
 ID VI7054 standard; cDNA; 1624 BP.  
 AC VI7054;  
 DT 08-JUN-1998 (first entry)  
 DE Glucosyl transferase (Gtase) encoding wound inducible gene (TW11).  
 KW Glucosyl transferase; Gtase; TW11; tomato; signalling pathway;  
 KW salicylic acid; jasmonic acid; ethylene; wound inducible gene;  
 KW plant defence protein; plant response; tobacco; rice; ss.  
 OS Lycopersicon sp.  
 FH Key Location/Qualifiers  
 FT CDS 2..1410  
 FT /\*tag= a  
 FT /product= "glucosyl transferase"  
 FT /note= "encodes Gtase from amino acid position 5  
 shown in W47172"  
 PN W09745546-A1.  
 PD 04-DEC-1997.  
 PF 30-MAY-1997; G01473.  
 PR 31-MAY-1996; GB-011420.  
 PA (UYVO-) UNIV YORK.  
 PI Bowles DJ, Calvert CM, O'Donnell PJ, Roberts MR;  
 DR WPI: 98-032653/03.  
 DR P-PSDB; W47172.  
 PT Tomato wound inducible (TW11) gene encoding glucosyl transferase -  
 PT useful to develop products that alter signalling pathways in plants

PT by altering of salicylic acid, jasmonic acid or ethylene  
 PS Claim 1: Fig 1; 52pp; English.  
 CC This wound inducible gene (TW11) isolated from wounded tomatoes encodes  
 CC a glucosyl transferase (Gtase) protein. This TW11 gene can be used to  
 CC identify homologue Gtase encoding genes isolated from tobacco and rice.  
 CC A microbial host can be transfected or transformed with a vector  
 CC containing the Gtase encoding nucleic acids. The products can be used to  
 CC interfere with Gtase and therefore alter signalling pathways in plants,  
 CC specifically tobacco, rice or tomato plants by altering levels of  
 CC salicylic acid, jasmonic acid or ethylene. This can induce the  
 CC production of plant defence proteins such as pathogenesis-related (PR)  
 CC and proteinase inhibitor (PIN) proteins which regulate plant development  
 CC (plant growth, reproduction and senescence) and improve plant response to  
 CC pathogens.  
 SQ Sequence 1624 BP; 545 A; 252 C; 372 G; 455 T;

Query Match 4.7%; Score 70; DB 1; Length 1624;  
 Best Local Similarity 49.8%; Pred. No. 2.6e-09;  
 Matches 221; Conservative 0; Mismatches 205; Indels 18; Gaps 1;

QY 791 gggggaataaactcgcgtgagtggtgaactcgaagccgaatacttcggtggtggtg 850  
 DB 779 GATGAACACGCGTGTGAATGGCTTGATTCGAAGAAATCAAGTTCCATTGTTATGTT 838

QY 851 tcgtttggagcgttttggaggtttccaaaggccacaaatggaagagattgggaaaggccta 910  
 DB 839 TGTTTTGGGAAGTACACACAGATTTCACTACAGACATGCACAGATTCGTTATGGGGCTA 898

QY 911 ttgacctgcggagggccctttttatgtagatgacgagacagaagaatgacgacggcgaa 970  
 DB 899 GAAGCCTCTGGCAAGATTTTCATTTGGTATTATCAGAACACAGGAATCAAGTTGGCTCCCA 958

QY 971 gaagaagaagaagaagaagattgagttgcattggggaattgaaaaaatgggaaaaata 1030  
 DB 959 GAAGGATTTCGAGGAAGAAAGAAC-----AAAAGAAAAGAGTTTAAATCATATA 1000

QY 1031 gtcgtcgtgctgcagtggtgaggttctggcgaccctgcgttgggagtttcgtgacg 1090  
 DB 1001 AGAGGATGGCACCCCAAGCTGTGATTCCTGATCAGCAAGCTATTGGAGCTTTTGTACT 1060

QY 1091 cattgtgggtgaaactcggctgtggagagcttgagttgcgggattccggtggtggtg 1150  
 DB 1061 CATTGTGGATGGAACCTCGACACTGGAAGGAATATCAGCAGGGGTACCAATGGTGACATGG 1120

QY 1151 ccgcagtggttgatcagcagcagaatggaagctgagagatcggtggtgggacaggg 1210  
 DB 1121 CCAGTATTTGCGGACACAGCTTTTTCATGAGAGAGTTGGTACTGAGGTAATGAGAAGTGGA 1180

QY 1211 gtgagagtgagaatgaatgaaggg 1234  
 DB 1181 GCTGGTGTGGTCTTAAGCAATGG 1204

RESULT 12  
 X24873  
 ID X24873 standard; cDNA; 1626 BP.  
 AC X24873;  
 DT 21-JUN-1999 (first entry)  
 DE Medium-chain UDP-glucose:fatty acid glucosyltransferase cDNA.  
 KW UDP-glucose:fatty acid glucosyltransferase; tomato;  
 KW 1-O-acyl-a-glucose; acyl donor; esterification; ss.  
 OS Lycopersicon pennellii.  
 OS Lycopersicon esculentum.  
 FH Key Location/Qualifiers  
 FT CDS 1..1416  
 FT /\*tag= a  
 PN W09909144-A1.  
 PD 25-FEB-1999.  
 PF 10-AUG-1998; IB1369.  
 PR 30-JUN-1998; US-106464.  
 PR 13-AUG-1997; US-055554.  
 PA (CORR ) CORNELL RES FOUND INC.



PA (PLBZ ) PLANT GENETIC SYSTEMS NV.  
PI Grootwassink JWD, Hemmingsen SM, Kolenovsky AD, Peferoen M;  
PI Reed DW, Underhill EW, Van Audenhove K;  
DR WPI: 97-247418/23.  
PT Plants genetically transformed to interfere with  
PT UDP-glucose:thiohydroximate S-glucosyltransferase gene expression  
PT - useful for production or rapeseed oil with reduced glucosinolate  
PT content  
PS Example 2: Page 20-21; 35pp; English.  
CC Brassica napus UDP-glucose:thiohydroximate S-glucosyltransferase  
CC (S-GT) partial cDNA clones pGL3-22 (T66178) and pGL4-2 (T66179)  
CC were obt'd. by PCR-RACE (see also T66175-77) of B. napus cDNA.  
CC The PCR product in pGL4 is completely contained in that of pGL3.  
CC pGL3 includes 982 bp of s-gt open reading frame, of which the last  
CC 116 bp overlap the g-gt fragment cloned in pGL2 (see also T66173-  
CC 74). A full-length cDNA clone (T66166) for B. napus S-GT (W09825)  
CC was subsequently obt'd.  
SQ Sequence 940 BP; 209 A; 285 C; 209 G; 237 T;  
Query Match 4.3%; Score 63.8; DB 1; Length 940;  
Best Local Similarity 45.4%; Pred. No. 9.2e-08;  
Matches 402; Conservative 0; Mismatches 462; Indels 21; Gaps 4;  
QY 74 cacataatcccgccctcccaattccgaagagactcctctaaagccgagcactgacgtcacg 133  
DB 1 CACCTCAACCAATGTCAGTTCGCTAAACGCTAGTCTCCAAAGCGCTCAAAGTTCACA 60  
QY 134 tttttcaagagcgtttatgcatggcgccgcatggccaaacacagcctccgcctgcgcga 193  
DB 61 ATCGCCACCACCACCTACACCGCTCTCTCATCTCCACCCCTCCCTCCGTCGAACCA 120  
QY 194 aaccacccggcctgacttcgttggcgcttcctcgagcgttagacgacgagggctgaagcc 253  
DB 121 ATCTCGAGCGGCACCATCTCATCTCCATCTCCACCCCTCCCTCCGTCGAACCA 180  
QY 254 ggcggcgagcgggaagcgttacatgtccgagatgaagccgcgcgcctccgagccttaaga 313  
DB 181 TCCGAATCTTCAAGTCTCAGCGCTCCCAACCTTAACCGCGGTATCTCCAAATTCMAA 240  
QY 314 aacctctctcaaacagcagcagcgcacatcttctgtcttacttccacactctttgcatg 373  
DB 241 TCACAGATTCGCCCATCGA- - - - -TTCTTTAGTCTACGACATCTTCTCCCGGTG 291  
QY 374 ggcggcgagcgttggcgcttgcacgctccgagcgcctctctctggtcgagccgc 433  
DB 292 GGAATCGAGTCGCGAGATCCAACTCCCTCTCAGCTGCGCGCTTCTTCACAAACCTC 351  
QY 434 accgtgctgtg---catataccacttctacttcaacggctcagcagcagatcgacgc 490  
DB 352 ACGTTTGTCTGTCTCTTCGCAATTCGCTCGGCTCGGCTGAGTTCTCTCCCGGTGATCC 411  
QY 491 ggttccaatgaaattcagctccctcggtctccatccctggagcagcagcttctccgag 550  
DB 412 GCTTCCGCGCTGTATCTGCTCGTGTGCTGCTGCGGCTTTCAGCTACGACGCTTCCTCC 471  
QY 551 ttctctgctcctgc 610  
DB 472 TTCGTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 531  
QY 611 ttgacggttgaagagaagcgaatattgttggaacacgtttgatcgcttggagcccgca- 669  
DB 532 ATTAACCATGAAGATGCTGATTGGCTGTTGCTCAACGGCTTTCGAAGGTTAGAGACACA 591  
QY 670 -----tgcactcaacgctattgataggtatgatgtgatcgagcgttgatccccc 724  
DB 592 GGTGTGAAGTTGGAGAAATCAGAGACTATGAAGCGCACGCTTTCATCGGACCTATGATCCA 651  
QY 725 tcgcctctcttgacgc 784  
DB 652 TCTGCTTATCTTGACGCCCGCAATCAAGACGATGAAGGCTACGGCTCAGCTGATG- - - 708  
QY 785 aaatcggaggagaataactcgtggagtggttgaaactcgaagccgcaaatcttcggtggtg 844

DB 709 AAGCCGCTCTGGGAGGAGTGTATGGAGTGTATAGACACTAAGCTAGTAAAGTCGGTGGT 768  
QY 845 tatgtgctgtttgggagcgtttttgaggtttccaaagccacaaatggaagagattgggaaa 904  
DB 769 TTTGTTTCGTTTGGTTCCTTTAGGATCCCTTTTGAAGCACTAGCTAGGTAGCAACG 828  
QY 905 ggcctattagcctgcggaagcgcctttttatgatgatacagaaa 949  
DB 829 GCGTTACAAGATCCAACTTTAACTTCTTGTGGTGTATTAAGAA 873  
RESULT 15  
Q74684  
ID 074684 standard; cDNA; 1650 BP.  
AC 074684; 1995 (first entry)  
DT 22-JUN-1995  
DE Early Ripening Tomato protein ERT1b gene.  
KW Early Ripening Tomato; transgenic plant; crop improvement;  
KW fruit ripening; transformation; ss.  
OS Lycopersicon esculentum.  
PN W09421794-A.  
PD 29-SEP-1994.  
PF 22-MAR-1994; G00581.  
PR 22-MAR-1993; GB-005860.  
PR 22-MAR-1993; GB-005862.  
PR 22-MAR-1993; GB-005865.  
PR 22-MAR-1993; GB-005859.  
PR 22-MAR-1993; GB-005859.  
PR 22-MAR-1993; GB-005866.  
PR 22-MAR-1993; GB-005869.  
PR 22-MAR-1993; GB-005867.  
PR 22-MAR-1993; GB-005868.  
PR 12-JUL-1993; GB-014351.  
PR 12-OCT-1993; GB-020988.  
PA (ZENEA) ZENEA LTD.  
PI Abu-Bakar UK, Barton SL, Gallego-Verigas PP, Gray JE;  
PI Grison D, Lowe AL, Picton S, Whotton LC, Abubakar UK;  
DR WPI: 94-317014/39.  
PT DNA constructs encoding fruit-ripening related proteins - useful  
PT for transformation of plants to modify fruit quality  
PS Claim 1: Page 47; 74pp; English.  
CC ERT1b homologous transcripts are 1.8 kb in size and it is only  
CC expressed during ripening of the wild-type fruit. Its expression is  
CC highest during early stages of fruit ripening (e.g. breaker plus 3).  
CC Levels of expression of the mRNA are low in rin (ripening inhibitor)  
CC mutant tomatoes and are restricted to ripening rin tomatoes. The  
CC gene is not activating upon wounding. The sequence has been  
CC deposited as NCIMB 40544. The DNA may be used to transform plants  
CC for production of plants with an altered level of expression.  
CC Improvements in transgenic plants include resistance to damage and  
CC pests, longer shelf life, improved flavor/aroma, etc.  
SQ Sequence 1650 BP; 492 A; 338 C; 348 G; 472 T;

Query Match 4.0%; Score 58.6; DB 1; Length 1650;  
Best Local Similarity 48.6%; Pred. No. 2.8e-06;  
Matches 191; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 791 gaggagaataactcgtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 850  
DB 732 GAAGACAATAGCTAGCTGCGAGTGGCTCGATACACAGCCCTTAACCTCTCTATGTC 791  
QY 851 tcgttggagcgttttgaggtttccaaaggccacaaatggaagagattgggaaaggccta 910  
DB 792 AGCTTCGGTAGCCTACTGAGGATTGATCACAAGAGTTGATTGAGACTGCTTGGGATTA 851  
QY 911 ttgacctgggaaggccctttttatgatgatacagacagagaagaatgacgcgcgcgaa 970  
DB 852 GCTAATAGCATCAACCGTTCTTGTGGTATTTCG- - -ACCTGGCTCTCTCTGCTTT 908  
QY 971 gaagaagaagaagaagaagaagttgagttgagttgagttgagttgagttgagttgagttg 1030  
DB 909 CAATGTGCTGAGGCACTGCCTGATGTTTGTGAGAAATGTTAGGAAATGTTAGGAAAGAGGACCAATA 968



**THIS PAGE BLANK (uspro)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:14:30 ; Search time 152.01 Seconds  
(without alignments)  
1333.842 Million cell updates/sec

Title: US-09-147-955-3  
Perfect score: 1474  
Sequence: 1 accaaacacaaacaaattt.....acttaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121.6	8.2	1731	4	US-08-466-583-1
2	121.6	8.2	1731	6	PCT-US95-07820-1
3	84	5.7	7218	1	US-08-232-463-14
4	70	4.7	1627	5	US-09-106-464-1
5	57	3.9	1669	3	US-08-522-421-1
6	51.2	3.5	1738	3	US-08-379-482A-2
7	49.8	3.4	1607	4	US-08-797-226-1
8	48.6	3.3	405	4	US-08-975-316-32
9	48	3.3	28958	1	US-08-258-261B-6
10	48	3.3	28958	1	US-08-456-837-6
11	48	3.3	28958	1	US-08-457-342-6
12	48	3.3	28958	1	US-08-457-646A-6
13	48	3.3	28958	2	US-08-458-076A-6
14	48	3.3	28958	2	US-08-764-233A-4
15	48	3.3	28958	2	US-08-457-335A-6
16	48	3.3	28958	2	US-08-729-214-6
17	48	3.3	49377	2	US-08-764-233A-1
18	47.6	3.2	2064	1	US-08-343-428-1
19	43.8	3.0	13987	1	US-08-804-227C-13
20	43.8	3.0	44377	3	US-08-804-227C-7
21	43.8	3.0	44377	4	US-08-804-198-1
22	42.2	2.9	2589	4	US-08-482-728A-3
23	41.4	2.8	1734	7	5352575-8
24	41.4	2.8	2336	6	PCT-US92-00282-1
25	41.4	2.8	2339	6	PCT-US92-00282-2
26	40.8	2.8	30001	1	US-08-125-468-1

c	27	40.8	2.8	30001	3	US-08-474-933-1	Sequence 1, Appli
	28	40.2	2.7	43280	3	US-08-804-227C-1	Sequence 1, Appli
	29	39.4	2.7	450	4	US-08-387-942C-40	Sequence 40, Appli
	30	39.4	2.7	12588	4	US-08-387-942C-1	Sequence 1, Appli
	31	39.4	2.7	53526	5	US-08-658-136-2	Sequence 2, Appli
	32	39.4	2.7	53577	5	US-08-658-136-1	Sequence 1, Appli
	33	38.2	2.6	2214	5	US-08-864-038A-1	Sequence 1, Appli
	34	38.2	2.6	3331	5	US-08-864-038A-2	Sequence 1, Appli
	35	38.2	2.6	3331	5	US-08-864-038A-4	Sequence 2, Appli
	36	38	2.6	1358	2	US-08-471-033-45	Sequence 4, Appli
	37	38	2.6	1358	3	US-08-471-044-45	Sequence 45, Appli
	38	38	2.6	1358	3	US-08-463-483A-45	Sequence 45, Appli
	39	38	2.6	1358	3	US-08-471-046A-45	Sequence 45, Appli
	40	38	2.6	1358	3	US-08-470-566B-45	Sequence 45, Appli
	41	38	2.6	1358	4	US-08-469-334-45	Sequence 45, Appli
	42	38	2.6	1358	5	US-09-300-529-45	Sequence 45, Appli
	43	37.4	2.5	6611	1	US-08-402-282-2	Sequence 2, Appli
	44	37.4	2.5	6611	1	US-08-508-004-2	Sequence 2, Appli
	45	37.4	2.5	6611	1	US-08-402-066-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-08-466-583-1  
; Sequence 1, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szczygowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: and Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,583  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1731 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 57..1472  
US-08-466-583-1

```

Query Match      8.2%; Score 121.6; DB 4; Length 1731;
Best Local Similarity 52.8%; Pred. No. 2.1e-24;
Matches 321; Conservative 0; Mismatches 269; Indels 18; Gaps 2;

QY 793 ggagaataactcgtgagtggttgaaactcgaagcgaataatcttcggtggtatgtgc 852
   ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 848 GGAGGATGCGTGACCAAGTGGCTAGACACCAAGCCGACCGCTCGTGCGCTACGTC 907

QY 853 gtttggagcgttttggagtttccaaagcacaataatggaagagatggaaggggtatt 912
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 908 CTTGCGAGCGCTCGCTCCCTGGGCAACGCCAGAGAGAGAGCTCGCGCGCGCTCCT 967

QY 913 agcctcggaagcctcttttattgagatgacagaaacagaagaatgacgacgagcgaaga 972
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 968 CGCGCGCGCAAGCCGCTCTGTTGGGTGGTAGGGC-----CAGCGACGA 1012

QY 973 agagaagaagaagagtggttgatgcttggggaattgaaataatggggaataatg 1032
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1013 GCACAGGTCCCGCGCTATCTCTCGGAGGCGACGCGCGCGCGATGCTGCT 1072

QY 1033 gtctggtgctcagtgagtggttctggcgacccctggttgggagtggttcgtgacgca 1092
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1073 GCCCTGTTGGAACTCCACGCTGGAGCGCTCAGCTTCGCGCTGCTTCGTACACCA 1132

QY 1093 ttgtgggtggaactcgtggttgagagcttgagttgaggagattcccggtggtggtgccc 1152
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1133 CTGCGGTTGGAACTCCACGCTGGAGCGCTCAGCTTCGCGCTGCTTCGTACACCA 1192

QY 1153 gcaagtgtttgacacacacgaatcgaaagctgattgagagatgcatgggagacaggggt 1212
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1193 GCTGTGACGACCGACCGCGGCGGAGCGGCGCTCGGAACGCTCGCTGGGCGCGGCGCT 1252

QY 1213 gagatgagaatgaatgaagg---gggtgggttgatggtgagatgagaagaagggtgtgt 1269
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1253 GCAGCGCGCGCGGATGCTGCGCGCGCGGCTGTTCTTCGGGGGGAAGTGGAGCGGTGCT 1312

QY 1270 ggagatggtgagtgatggagggggtgacaaagacaaactagttagaggaanaatgcctcaaatg 1329
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1313 GCAGCGCGCTCATGGACGGGGGCGAGCGGCGCTCTGTCACGCAAGCGCGCGGGAATG 1372

QY 1330 gaagactttggccagacagaagcctatggatgagatgattcttcaactcaaatctcaacgc 1389
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1373 GAGGACAGGCTCGCGCGCGGCTGACACCGCTGCGAGCTCTGACCGCAACCTGGACGA 1432

QY 1390 cttttcttc 1397
   |||||
Db 1433 GTTCGTGC 1440

```

```

RESULT 2
PCT-US95-07820-1
; Sequence 1, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Farber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1731 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 57..1472
; PCT-US95-07820-1

```

```

Query Match      8.2%; Score 121.6; DB 6; Length 1731;
Best Local Similarity 52.8%; Pred. No. 2.1e-24;
Matches 321; Conservative 0; Mismatches 269; Indels 18; Gaps 2;

QY 793 ggagaataactcgtgagtggttgaaactcgaagcgaataatcttcggtggtatgtgc 852
   ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 848 GGAGGATGCGTGACCAAGTGGCTAGACACCAAGCCGACCGCTCGTGCGCTACGTC 907

QY 853 gtttggagcgttttggagtttccaaagcacaataatggaagagatggaaggggtatt 912
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 908 CTTGCGAGCGCTCGCTCCCTGGGCAACGCCAGAGAGAGAGCTCGCGCGCGCTCCT 967

QY 913 agcctcggaagcctcttttattgagatgacagaaacagaagaatgacgacgagcgaaga 972
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 968 CGCGCGCGCAAGCCGCTCTGTTGGGTGGTAGGGC-----CAGCGACGA 1012

QY 973 agagaagaagaagagtggttgatgcttggggaattgaaataatggggaataatg 1032
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1013 GCACAGGTCCCGCGCTATCTCTGCGGAGCGACGCGCGCGCGATGCTGCT 1072

QY 1033 gtctggtgctcagtgagtggttctggcgacccctggttgggagtggttcgtgacgca 1092
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1073 GCCCTGTTGGAACTCCACGCTGGAGCGCTCAGCTTCGCGCTGCTTCGTACACCA 1132

QY 1093 ttgtgggtggaactcgtggttgagagcttgagttgaggagattcccggtggtggtgccc 1152
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1133 CTGCGGTTGGAACTCCACGCTGGAGCGCTCAGCTTCGCGCTGCTTCGTACACCA 1192

QY 1153 gcaagtgtttgacacacgaatcgaaagctgattgagagatgcatgggagacaggggt 1212
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1193 GCTGTGACGACCGACCGCGGCGGAGCGGCGCTCGGAACGCTCGCTGGGCGCGGCGCT 1252

QY 1213 gagatgagaatgaatgaagg---gggtgggttgatggtgagatgagaagaagggtgtgt 1269
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1253 GCAGCGCGCGCGGATGCTGCGCGCGGCTGTTCTTCGGGGGGAAGTGGAGCGGTGCT 1312

QY 1270 ggagatggtgagtgatggagggggtgacaaagacaaactagttagagagaataatgcctcaaatg 1329
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1313 GCAGCGCGCTCATGGACGGGGGCGAGCGGCGCTCTGTCACGCAAGCGCGCGGGAATG 1372

QY 1330 gaagactttggccagacagaagcctatggatgagatgattcttcaactcaaatctcaacgc 1389
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1373 GAGGACAGGCTCGCGCGCGGCTGACACCGCTGCGAGCTCTGACCGCAACCTGGACGA 1432

QY 1390 cttttcttc 1397
   |||||
Db 1433 GTTCGTGC 1440

```



QY	791	gaggagaataactcgtagtggttgaaactgaagcgaatacttcggtggtgtagtg	850
Db	793	GATGAACACGCGTGCTGGAATGGCTTGATTCGAAGAAATCAAGTTCCTATGTT	852
QY	851	tcatttggagacgcttttgaagtttccaaagacacaatgaagagattgggaagagagcta	910







```
;
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-258-261B-6

Query Match          3.3%; Score 48; DB 1; Length 28958;
Best Local Similarity 46.4%; Pred. No. 0.0022;
Matches 156; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 156 gggcgccgcatggcacaacagctccgctccgctccggaacccaccgggctcgacttcg 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10208 GCCACAGCCGAGATCGCCGCTTCGTCGAGGCGCTCTCTCCCTCGAGGACG 10267

QY 216 tggcgttctccgacggtacgacagcgggctgaagcccgcgacggaagcgctaca 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10268 CGGCGCCGATCGCCGCTTCGCGAGGAAAGCGCTCACACCGTGGCGGCAACGGCGGCA 10327

QY 276 tgtccgagatgaagccgctccgctccgcttaagaacacctcttctcaaacagcagc 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10328 TGGCGCGCTCGAGCTCGCGCTTCGACCTCCAGACTTACTCTCCCTGGGGCGACA 10387

QY 336 acgtcaactttgctgtactccacctcttgcattggcgaggagtggtggtttgt 395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10388 GGCTCTCACCGCGCGCTCAACAGCCCGAGGCTACCTCGTATCGCGGAGCCCGCG 10447

QY 396 cccacgtcccgacgcctctcttgggtcgagcccgccacccgctgtgtatataccact 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10448 CGGTGAGCGGCTCTCGAGCTCTCACCGCCACCAAGGTGTTGCGCCCGCAAGATCGCG 10507

QY 456 tctactcaacgggtacgacgagatcgacgcg 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10508 TCGACTACGCTCCCACTCCGCCAGATGGAGCGCG 10543

RESULT 10
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-456-837-6

Query Match          3.3%; Score 48; DB 1; Length 28958;
Best Local Similarity 46.4%; Pred. No. 0.0022;
Matches 156; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 156 gggcgccgcatggcacaacagctccgctccgctccggaacccaccgggctcgacttcg 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10208 GCCACAGCCGAGATCGCCGCTTCGTCGAGGCGCTCTCTCCCTCGAGGACG 10267

QY 216 tggcgttctccgacggtacgacagcgggctgaagcccgcgacggaagcgctaca 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10268 CGGCGCGATCGCCGCTTCGCGAGGAAAGCGCTCACACCGTGGCGGCAACGGCGGCA 10327

QY 276 tgtccgagatgaagccgctccgctccgcttaagaacacctcttctcaaacagcagc 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10328 TGGCGCGCTCGAGCTCGCGCTTCGACCTCCAGACTTACTCTCCCTGGGGCGACA 10387

QY 336 acgtcaactttgctgtactccacctcttgcattggcgaggagtggtggtttgt 395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10388 GGCTCTCACCGCGCGCTCAACAGCCCGAGGCTACCTCGTATCGCGGAGCCCGCG 10447

QY 396 cccacgtcccgacgcctctcttgggtcgagcccgccacccgctgtgtatataccact 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10448 CGGTGAGCGGCTCTCGAGCTCTCACCGCCACCAAGGTGTTGCGCCCGCAAGATCGCG 10507

QY 456 tctactcaacgggtacgacgagatcgacgcg 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10508 TCGACTACGCTCCCACTCCGCCAGATGGAGCGCG 10543

RESULT 11
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
```



Db 10388 GGCTCTCCAGCGCGCTCAACAGCCCCAGGGCTACCTCGTATCCGCGGAGCCCGCG 10447  
QY 396 ccaagtcacagcgcctctctctggtgagccgcacccgctgtgtgatatataccact 455  
Db 10448 CCGTCGAGCGGCTGCTCGAGGCTCTCACCGCCACCAAGGTGTTCCGCCGCAAGATCCGGC 10507  
QY 456 tctactcaacggctacgacgagatcgacgcg 491  
Db 10508 TCGACTACGCTCCCACTCCGCCAGATGGACGCCG 10543

## RESULT 13

US-08-458-076A-6

; Sequence 6, Application US/08458076A

; Patent No. 5698425

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James M.

; APPLICANT: Beck, James Joseph

; APPLICANT: Hill, Dwight Steven

; APPLICANT: Ryals, John Andrew

; APPLICANT: Gaffney, Thomas Deane

; APPLICANT: Lam, Stephen Ting

; APPLICANT: Hammer, Phillip E.

; APPLICANT: Uknes, Scott Joseph

; TITLE OF INVENTION: Genes for the synthesis of

; TITLE OF INVENTION: antipathogenic substances

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ciba-Geigy Corporation

; STREET: 7 Skyline Drive

; CITY: Hawthorne

; STATE: NY

; COUNTRY: USA

; ZIP: 10532

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/458,076A

; FILING DATE: 01-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/457,205

; FILING DATE: 01-JUN-1995

; APPLICATION NUMBER: 08/258,261

; FILING DATE: 08-Jun-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Elmer, James Scott

; REGISTRATION NUMBER: 36,129

; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3

; TELEPHONE: 919-541-8614

; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 28958 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-458-076A-6

Query Match

Best Local Similarity 3.3%; Score 48; DB 2; Length 28958;

Matches 156; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 156 ggcgcgcagtgccacacagcctccgcgctgcggaacccacccggcctcactcg 215  
Db 10208 GCCACAGCCAGGCGAGATCGCCCGCCTTCGTCGAGGCGCTCTTCCCTCGAGGACG 10267  
QY 216 tggcgtctccgcagcgtctacgacgacgggctgaagcccgcgcgagcggaagcgctaca 275  
Db 10268 CGGCGCGCATCGCGCGCCTGCGCAGGAAGGCTCACACACGTCGGCGGCAACGCGGCA 10327  
QY 276 tgcgagatgaagcccgcggtccgagccttaagaaacctctcttctataaacgacg 335  
Db 10328 TGGCCGCGCTCGAGATCGGCGCCTCCGACCTCCAGACCTACTCGCTCCCTGGGCGGACA 10387  
QY 336 acgtcactttcgctctactctccacctctttgcatggcgcgaggtggcgctttgt 395  
Db 10388 GGCTCTCACCGCGCGCTCAACAGCCCCAGGGCTACCTCGTATCCGCGGAGCGCCGCG 10447  
QY 396 ccaagtcacagcgcctctctctggtcgagcccgccacccgctgtgtgatatataccact 455  
Db 10448 CCGTCGAGCGGCTGCTCGACGCTCTCACCGCCACCAAGGTGTTCCGCCGCAAGATCCGCG 10507  
QY 456 tctactcaacggctacgacgagatcgacgcg 491  
Db 10508 TCGACTACGCTCCCACTCCGCCAGATGGACGCCG 10543

## RESULT 14

US-08-764-233A-4

; Sequence 4, Application US/08764233A

; Patent No. 5716849

; GENERAL INFORMATION:

; APPLICANT: Ligon, James M.

; APPLICANT: Schupp, Thomas

; APPLICANT: Beck, James J.

; APPLICANT: Hill, Dwight S.

; APPLICANT: Neff, Shezanna

; APPLICANT: Ryals, John A.

; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ciba-Geigy Corporation

; STREET: 520 White Plains Road, P.O. Box 2005

; CITY: Tarrytown

; STATE: NY

; COUNTRY: USA

; ZIP: 10591

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/764,233A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/729,214

; FILING DATE: 09-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/258,261

; FILING DATE: 08-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Meigs, J. Timothy

; REGISTRATION NUMBER: 38,241

; REFERENCE/DOCKET NUMBER: 1506/CIP6

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (919) 541-8587

; TELEFAX: (919) 541-8689

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 28958 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 18:20:52 ; Search time 51.03 Seconds  
(without alignments)  
557.848 Million cell updates/sec

Title: US-09-147-955-2  
Perfect score: 2421  
Sequence: 1 MVRRRVLLATFPAQGHINPA.....GEDGSLKLNLAFLHQVARA 460

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_64: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1053.5	43.5	467	2	T02238	glucosyl transferase
2	1038	42.9	456	2	C71402	probable glucosylt
3	888	36.7	458	2	C71420	hypothetical prote
4	762.5	31.5	471	2	A54739	indole-3-acetate b
5	702	29.0	479	2	E71419	probable indole-3-
6	675.5	27.9	460	2	T00639	hypothetical prote
7	663.5	27.4	456	2	T00506	indole-3-acetate b
8	658.5	27.2	484	2	D71419	probable indole-3-
9	653.5	27.0	438	2	T00507	indole-3-acetate b
10	628.5	26.0	475	2	F71419	probable indole-3-
11	604.5	25.0	453	2	T00511	indole-3-acetate b
12	501.5	20.7	453	2	T45603	glucosyltransferas
13	484	20.0	455	2	T00584	indole-3-acetate b
14	466.5	19.3	451	2	T45604	glucosyltransferas
15	458.5	18.9	452	2	T12981	hypothetical prote
16	455.5	18.8	440	2	T00583	indole-3-acetate b
17	455	18.8	438	2	T45602	glucosyltransferas
18	445	18.4	449	2	T45605	glucosyltransferas
19	444.5	18.4	447	2	T12978	hypothetical prote
20	414.5	17.1	462	2	T01732	UTP-glucose glucos
21	413	17.1	347	2	T06371	probable UDP-glucu
22	407.5	16.8	452	2	T00981	flavonol 3-O-glucu
23	406.5	16.8	466	2	T07404	flavonol 3-O-glucu
24	406.5	16.8	471	2	S01037	flavonol 3-O-glucu
25	403.5	16.7	385	2	T12980	hypothetical prote
26	400.5	16.5	476	2	T03745	glucosyltransferas
27	398.5	16.5	420	2	T08005	flavonol 3-O-glucu
28	398	16.4	433	2	S51767	glucosyl transfera
29	398	16.4	452	2	G71416	probable glucosylt

## ALIGNMENTS

RESULT 1

T02238

glucosyl transferase, jasmonate-induced - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 24-Nov-1999

C:Accession: T02238

R:Kojima, H.; Hashizume, K.; Imanishi, S.; Nakamura, K.

A:Submitted to the EMBL Data Library, January 1997

A:Description: Jasmonate-induced potential glucosyltransferase from tobacco suspension

A:Reference number: Z14633

A:Accession: T02238

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-467 &lt;KOU&gt;

A:Cross-references: EMBL:AB00623; NID:d1094897; PID:d1019901

A:Experimental source: strain BY-2

C:Genetics:

A:Gene: JIGT

C:Superfamily: flavonol O3-glucosyltransferase

Query Match	43.5%	Score	1053.5	DB	2	Length	467
Best Local Similarity	45.3%	Pred. No.	3.4e-77				
Matches	209	Conservative	83	Mismatches	150	Indels	19
Gaps							7
Qy	6	VLLATFPAGHINPALQFAKRLKAGTDVFTTSVYAVRRMANTASAAAGNPGLDFVAF	65				
Db	10	VLIALFPQGQHINPSLQFSKLLNLGVKVTLSLSLAFNRKIKLPKI-----EGLTFAPF	64				
Qy	66	SDGYDDGLKPCDGRK-YMSEMKARGSEALRNLL-----LNNHDVTFVYSHLFAWAAEYA	120				
Db	65	SDGYDGNFKGSDDDYHLFNSAIKSHGSEFIANLISKAKNGYPTFRVITYILMDWAGSA	124				
Qy	121	RSQVPSALLWPEATVLCIYFYFNGYAD---EIDAGSDEIQLPRLPLEQRSIPFTLL	177				
Db	125	KKLHPSTLFWIQPATVFDIYYRTNFANYFNKVDSDQIQLPGLPSLSSDFPSFVF	184				
Qy	178	PETPERFRILM--MKKLETLDGEEKAKVLVNTFDALPDALTADRIYELIGLIPISAF	235				
Db	185	DDVKSNDNAVESIKKQIETLSEENPRILVNTFDALNLALRVKLVNTVGVIGLIPISF	244				
Qy	236	LDGGPSETSYYGDLFPEKSENNCENVDLTKPKSSVYVYSGSVLPFPAQMEETGKGLL	295				
Db	245	LDEKDRKNFFAADIIE--SENNYMEWLDARANKSVIYTFAGSYAEISSQWMEETISQGLL	302				
Qy	296	ACGRFLFMIRKQKNDGDEEEELSCIGELKMGKIVSMCSQLEVLHAPALGCFVTHCG	355				
Db	303	KGRFLFWIRTLN--GEKPEKLTCKDELEKIGRIVRCQMEVKKHSSVGCFLTHCG	360				
Qy	356	WNSAVESLSCGVPVAVPQWFDQTTNAKLIEDAWGTGVRVRNNEGGVDSIERCVEMV	415				
Db	361	WNSTLESASGVPIVACPIWNDQICNAKLIQDWKIGVRVNAKKEGIIKRDQFQKIEIV	420				

30	397.5	16.4	471	1	S08325	flavonol 3-O-glucu
31	397.5	16.4	471	2	S01052	flavonol 3-O-glucu
32	396.5	16.4	478	2	T05423	probable glucosylt
33	395	16.3	472	2	S39507	glucuronosyl trans
34	392.5	16.2	476	2	T03747	glucosyltransferas
35	391	16.2	488	2	T07786	UDP-glucose glucos
36	389.5	16.1	455	1	XUBHFG	flavonol 3-O-glucu
37	384	15.9	487	2	S41951	UTP-glucose glucos
38	379.5	15.7	478	2	A71417	hypothetical prote
39	367.5	15.2	394	2	S41952	UTP-glucose glucos
40	367.5	15.2	478	2	T08395	glucosyltransferas
41	363	15.0	490	2	T46162	UTP-glucose glucos
42	361.5	14.9	346	2	S41954	UTP-glucose glucos
43	358.5	14.8	481	2	T01850	UTP-glucose glucos
44	357	14.7	287	2	S41953	glucosyltransferas
45	356.5	14.7	507	2	T46161	glucosyltransferas



A:Accession: A54739  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-471 <SE>  
A:Cross-references: GB:I34847; NID:g548194; PID:AAA59054.1; PID:g548195  
C:Superfamily: flavonol O3-glucosyltransferase  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 31.5%; Score 762.5; DB 2; Length 471;  
Best Local Similarity 39.1%; Pred. No. 9.9e-54;  
Matches 187; Conservative 70; Mismatches 182; Indels 39; Gaps 16;  
Qy 6 VLLATFFPAQHINPALQFAKRLKAGTDTVTFTSYAMRMANTASAAAGNPPGLDFVAF 65  
Db 5 VLVPFPQGHNPVQFAKRLKAGTDTVTFTSYAMRMANTASAAAGNPPGLDFVAF 65  
Qy 66 SDGVDGGLKPCGDG-KRYMSEMKARGSLRNLL---LNNHVDVTFVYVSHLFAWAEVA 120  
Db 58 SDGHDGEGFASAGVAEYLEKQAAASLASLVEARASSADAFCTCVYDSDVDWLPVA 117  
Qy 121 RESQVPSALLWEPATVLCIYFFENG-----YADEIDAGSDEIOLPR----LPLE 168  
Db 118 RMGLPAVPFSTQCAVSAVYTHFSQGRVAVPPGAADGSDGGAALSAEAFGLPEME 177  
Qy 169 ORSLPTFLPETPERFRLMKKEKLETDGEEKAKVLVNTFDALPDALTAIDRY-ELIGI 227  
Db 178 RSELSFVDFHGPYTIAMQAIKQFAHAGKDD-WVLFNSFELETEVLAKTKYLKARAI 236  
Qy 228 GPLI--PSAFLDGGDPSTSYGGDLFEKSEENNCVWLDTPKSSVVVYVSGSVLRFPKA 285  
Db 237 GPCVPLTAGRTAGANRITGANLVK--PEDACTKWLDTKPDRSVAVVYVSGSLASLGN 294  
Qy 286 QWEETGKGLLACGRPFLLMIREQKNDGDEEEELSCIGELKMG--KIVSWCSOLEVLA 343  
Db 295 QKEELARGLLAAGKPFLLVAVRA-----SDEHQVPRYLLAEATATGAAMVVPWCPLDVL 349  
Qy 344 IPALCGFTVHCWNSAVESLSCGVPVAVPQWFDQTTAKLIEDAWGTGVRVRNMGEGV 403  
Db 350 HPVACFTVHCWNSLALSGFVPMVAMALMDPTNARVNLANGAGVARRDAGAGV 409  
Qy 404 D-GSBIERCVMVMDGGEKSLVRENAIKWTLAREAMGEDGSSLLKLNLAFLHQVRA 460  
Db 410 FLRGVERCVRAVMDGGEAASAAKACGEWRDARAAPVAGSSDRNLDEFV-QPVRA 466  
RESULT 5  
E71419  
probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: columbia  
C:Accession: E71419  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; Giel  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; An  
C.; Chaiwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis tha  
A:Reference number: A71400; MUID:98121113  
A:Accession: E71419  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-479 <BE>  
A:Cross-references: GB:297339; NID:g2244901; PID:e326931; PID:g2244906  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 29.0%; Score 702; DB 2; Length 479;

Best Local Similarity 34.9%; Pred. No. 7.7e-49;  
Matches 166; Conservative 87; Mismatches 182; Indels 40; Gaps 14;  
Qy 6 VLLATFFPAQHINPALQFAKRLKAGTDTVTFTSYAM---RRMANTASAAAGNPPGLDF 62  
Db 9 VMLVFPQGHNPVQLRLGLKIASGLLVTFVTTEKPMGKKMRQANKIQDGLVKPVLGLGF 68  
Qy 63 VA---FSDGYDDGLKPCGDGKRYMSEMKARGSEALRNLL---LNNHVDVTFVYVSHLFAWAA 117  
Db 69 IRFEFSDGFADDDKRFDFDAFRPHLEAVGKQEIKNLVKRYNKEPVTCLINNAFVFWC 128  
Qy 118 EVARESQVPSALLWEPATVLCIYFFENGVA-----DEIDAGSDEIOLPRLPLEQSRSL 172  
Db 129 DVAEELHIPSAVLWYOSCACTAYVYVHHRLVKFKTPKTEPDI---SVEIPCLPLKHDDEI 185  
Qy 173 PTFLLPEP-ERFRLMKKEKLETDGEEKAKVLVNTFDALPDALTAIDRY-----ELIGI 227  
Db 186 PSFLHPSPYTAFGDIILDLQKRFENHKSFYLFIDTFRELEKDIMHMSQLCQPAIISPV 245  
Qy 228 GPLIPSAFLDGGDPSTSYGGDLFEKSEENNCVWLDTPKSSVVVYVSGSVLRFPKAQM 287  
Db 246 GPLFKMAQTLSDDVK-----GDISEPA--SDCMEWLDSEPSVYVYISFGTTANLKQEQM 298  
Qy 288 EEIGKGLLACGRPFLLMIREQKNDGDEEEELSCIGELKMGKIVSWCSQLEVLHAHPAL 347  
Db 299 EEIAGHVLSSGLSVLWVVRPPM--EGTFEPHV-LPRELEKGIKIVWCPOEVLHAHPAI 355  
Qy 348 GCFVTHCGWNSAVESLSCGVPVAVPQWFDQTTAKLIEDAWGTGVRVRNMGEGVDSGE 407  
Db 356 ACFLSHCGWNSMETALTAGVPVCPQDQVTDVAVLADVFKTGVRVRL---GRGAAEEM 411  
Qy 408 I---ERCVEMVMDG--GEKSLVRENAIKWTLAREAMGEDGSSLLKLNLAFLHQV 457  
Db 412 IVSRVVAEKLLEATVGEKAVELRENAIKWTLAREAMGEDGSSLLKLNLAFLHQV 466  
RESULT 6  
T00639  
hypothetical protein F3I6.2 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999  
C:Accession: T00639  
R:Vysotskaya, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.  
Submitted to the EMBL Data Library, February 1998  
A:Reference number: Z14197  
A:Accession: T00639  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-460 <FE>  
A:Cross-references: EMBL:AC002396; NID:g2749918; PID:g2829862; GSPDB:GN00059; ATSP:F3  
C:Genetics:  
A:Gene: ATSP:F3I6.2  
A:Map position: 1  
A:Introns: 219/3  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 27.9%; Score 675.5; DB 2; Length 460;  
Best Local Similarity 34.9%; Pred. No. 1e-46;  
Matches 169; Conservative 72; Mismatches 186; Indels 57; Gaps 15;  
Qy 2 VRRVLLATFFPAQHINPALQFAKRLKAGTDTVTFTSYAMRMANTASAAAGNPPGLD 61  
Db 8 VKGHVVLIPYPVQGHNPVQFAKRLKAGTDTVTFTSYAMRMANTASAAAGNPPGLD 58  
Qy 62 FVAESDGYDDGLKPCG---DGKRYMSEMKARGSEALRNLL---LNNHVDVTFVYVSHLF 113  
Db 59 VEPISDGF--FIPIGIFGFSVDITYSEFKLNGSETTLILLIEKFKSTSPIDCLISFL 116  
Qy 114 AWAEEVARESQVPSALLWEPATVLCIYFFENG-YADEIDAGSDEIOLPRLPLEQSRSL 172  
Db 117 PWGLEVARSMELSAASFFTNLTVCVSLKFSNGDFPLPADPNSAPFRIGLPSLSYDEL 176



erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;  
C.; Chawatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t  
A:Reference number: A71400; MUID:98121113  
A:Accession: F71419  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-475 <BEV>  
A:Cross-references: GB:Z97339; NID:g2244901; PID:e326932; PID:g2244907  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: flavonol O3-glucosyltransferase

## RESULT 9

T00507

Indole-3-acetate beta-glucosyltransferase homolog T20D16.12 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999

C:Accession: T00507

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul

submitted to the EMBL Data Library, November 1997

A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.

A:Reference number: Z14159

A:Accession: T00507

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-438 &lt;ROU&gt;

A:Cross-references: EMBL:AC002391; NID:g2642427; PID:g2642438

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 2

A:Note: T20D16.12

C:Superfamily: flavonol O3-glucosyltransferase

Query Match 27.0%; Score 653.5; DB 2; Length 438;

Best Local Similarity 35.1%; Pred. No. 5.5e-45;

Matches 160; Conservative 88; Mismatches 169; Indels 39; Gaps 17;

QY 14 QGHINPALQFAKRLKAGTDTFTTSVAVRRMANTASAAAGNPGLDFVAFSDGY--DD 71

Db 7 OGLNPMKFAKHLARTLHFTLATTEQARDLLSSTADEP--HRP-VDLAPFSDGLPKDD 63

QY 72 GLKPCGDKRYNSEKARGSEALNLLNNHDDTVFVYSHLFAWAAVARESQVPSALLW 131

Db 64 PRDP-----DTLAKSLKDGAKNL-SKIIEEKRFDCCIISVPFPWPVPAVAAAHNIPCAILW 118

QY 132 VEPATVLCIYYFYF---NGYADEIDAGSDEIQLPRLPLEQSLTFLPETPERFLMM 188

Db 119 IQACGAFSYYYRYMKNPFPDLEDL-NOTVELPALPLEVRDLPSLMLPSOGANVTILM 177

QY 189 KEKLTLDGEERAKVLNFTFDALPDALTAI--DRVELIGIGLPIPSAFIDGGDPSETSYG 247

Db 178 AEFADCL--KDVKWLVSFYSLESEIETESMSDLPIIPIGPLV-SPFLGNDDEKTL-- 232

QY 248 GDLFEKSENNCVENLDTKPKSVVYSGSVLRFPKQMEIGKGLACGRPFLLMIRE 307

Db 233 -DMWK--VDDYCMWLDQARSVYIISFGSILKSLNQVETIATLKNRGVFFLWVIRP 289

QY 308 OKNDGDEEEELSCIGELKKMGK--IVSWCSQLEVLALHPALGCFVTHCGWNSAVESLSC 365

Db 290 -----KEGENVQVLQENVKEGKVVTWEGQOEKILSHMAISCFITHCGWNSTIETVVT 343

QY 366 GVPVAVPQWFDQTTNAKLIEDAWGTGVRVRNNEGGVDG-----SEIERCVEMVMDGGEK 421

Db 344 GVPVAVPTWIDQPLDARLLVDVFGIVRMK---NDAIDGELKVAEVEICIEAVTEGPAA 400

QY 422 SKLVRENALKKTLAREANGDGSLSKLNAPLHQV 457

Db 401 ADM-RRRATLKHAAARSAMPGSSAQNLDSFISDI 435

## RESULT 10

F71419

probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

A:Variety: Columbia

C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999

C:Accession: F71419

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk

P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel

avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.

Nature 391, 485-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech

Query Match 26.0%; Score 628.5; DB 2; Length 475;

Best Local Similarity 33.3%; Pred. No. 6.4e-43;

Matches 159; Conservative 85; Mismatches 183; Indels 51; Gaps 17;

QY 6 VLLATFPAGQHINPALQFAKRLKAGTDTFTTS---VYAWRRMANTASAAAGNPGLDF 62

Db 10 VMLVSFPGGHSPLRLGKIITASKGLIVTFVTEEPGLCKMKRQANNIOGVLPVGLGF 69

QY 63 VAFSDGYDDGLKPCGDKRYNSEKARGSEALNLL--LNNHDDTVFVYSHLFAWAAVYA 120

Db 70 LRF-EFFEDGFYKEDFOLLQKSLVSGRKREIKNLVKYKQPVRCILINNAFVPMVCDIA 128

QY 121 RESQVPSALLWTEPATVLCIYYFYFNGYADEIDAGSDEI--QLPRLP-PLEORSLETFLL 177

Db 129 BELQIPSAVLWQSCACLAAYYYYHHQLVKFPTETETPEITVDVPFKPLTKHDEIFSFLL 188

QY 178 PTPPERFLMKKEKLETLDE--ERAK-----VLVNTFDALPDALTAID----RYEL 224

Db 189 PESP-----LSSIGGTILEQIKRLHKPFSLIETFOLEKTDIDHMSQLCPQVNF 238

QY 225 IGIGLPIPSAFIDGGDPSETSYGGDLFEKSENNCVENLDTKPKSVVYSGSVLRFPK 284

Db 239 NPIGLFTMA-----KTIKSDIKGDI--SKPDSDCIEMWLDSEKREPSSVYISFGLAFLQ 291

QY 285 AQMEELGKGLACGRPFLLMIREQKNDDGEEEEEELSCIGELKMKGIYSWCSQLEVLH 344

Db 292 NQIDEIAHGLNSGLSCLWLVRPL--EGLAIEPHVPL--ELEEKGIYEWQOEKVLH 348

QY 345 PALGCFVTHCGWNSAVESLSCGVPVAVPQWFDQTTNAKLIEDAWGTGVRVRNNEGGV-- 402

Db 349 PAVACFLSCHGWSNMEALTSGVPVICFQWGDVNTAVYMIDVFTG--LRLSRGASDE 406

QY 403 --VDGSEI--ERCVMYMDGGEKSLVRENAIKWTKLAREAMGEDGSSKLNAPLHQV 457

Db 407 RIVPREEVAERLEATV--GEKAVELRENARRWKEAESAVAYGGSERNFQEFVDKL 462

## RESULT 11

T00511

Indole-3-acetate beta-glucosyltransferase homolog T20D16.16 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999

C:Accession: T00511

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K

submitted to the EMBL Data Library, November 1997

A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.

A:Reference number: Z14159

A:Accession: T00511

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-453 &lt;ROU&gt;

A:Cross-references: EMBL:AC002391; NID:g2642427; PID:g2642442

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 2

A:Introns: 266/2

A:Note: T20D16.16

C:Superfamily: flavonol O3-glucosyltransferase

Query Match 25.0%; Score 604.5; DB 2; Length 453;  
 Best Local Similarity 33.1%; Pred. No. 5.2e-41;  
 Matches 155; Conservative 82; Mismatches 178; Indels 53; Gaps 15;

QY 6 VLLATPFAQGHINPALQFAKRLKAGDVDTFTTSVYAWRRMANTASAAAGNPPGL-DFVA 64  
 Db 11 VLVVALPFGHLNPLMKFAKHLARTNLHFTLATIESARDLSST-----DEPHSLVDLVF 65

QY 65 FSDGYDDGLPCGDKRYMSEMKARGSEALRNLLNHHVDYF-----VYSHLFAWAA 117  
 Db 66 FSDGLPK-----DDPRDHEPL-----TESLRKYGANNFSKIEGKRPDCIISVPFPPWP 115

QY 118 EVARESOPVALLWPEPATVLCIYFYFF---NGYADIDAGSDEIOLPRLPPELQSLPT 174  
 Db 116 AVAAAHNIPCAILWIEACAGFSVYRYMYKTNFSD-LEDPNKGVELPGLPFLEVRDLPT 174

QY 175 FLIPETPERFLMMKEKLETLGDEEKAKVLVNTFDALPDALTAI-DRYELIGIGPLIPS 233  
 Db 175 LMLPSHGAIFNTLMAEFVECL--KDKWVLANSFYELSVIESMFDLKPIIPIGPLV-S 231

QY 234 AFLDGGDPSETS YGGDLFEKSEENNCVWLDTPKSSVYVVSFGSVLRFPPKAQMEIEGKG 293  
 Db 232 PELGADDEKILDGKSLDMWKADDCHEWLDQV-----SILKSENQVETIATA 281

QY 294 LLACGRPFLMIREQKNDGDEEEELSCIGELKMGKIYSCWQSLEVLHAPALGCFVTH 353  
 Db 282 LKNRGVPFLAVIRPKEAENVVLEDAVESQ---GVVIENGQOEKILCHMAISCFVTH 337

QY 354 CGNVSATESUSCGVPVAVPQWFDQTTNAKLIEDAWGTQVVRVMEGGVDG-----SEIE 409  
 Db 338 CGNSTIETVSGVPVAVPTWFDQPLDARLLVDVFGIVRMKNDV---VDGELKVAEVE 394

QY 410 RCVMWMDGGEKSLVRENAIKWKTLEAREAMGEGSSLKLNALFLHOV 457  
 Db 395 RCIDAVTKGTDAADM-RRRAELKQATRSAMAPCGSLARNLDLFINDI 441

RESULT 12  
 T45603  
 glucosyltransferase-like protein - Arabidopsis thaliana  
 N;Alternate names: protein F12A12.180  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000  
 C;Accession: T45603  
 R;Choi, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa  
 submitted to the Protein Sequence Database, December 1999  
 A;Reference number: Z23008  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-453 <CHO>  
 A;Cross-references: EMBL:AL133314  
 A;Experimental source: cultivar Columbia; BAC clone F12A12  
 C;Genetics:  
 A;Map position: 3  
 A;Note: F12A12.180  
 C;Superfamily: flavonol O3-glucosyltransferase

Query Match 20.7%; Score 501.5; DB 2; Length 453;  
 Best Local Similarity 29.5%; Pred. No. 1e-32;  
 Matches 142; Conservative 78; Mismatches 193; Indels 69; Gaps 14;

QY 3 RRRVLLATPFAQGHINPALQFAKRLKAGDVDTFTTSVYAWRRMANTASAAAGNPPGLDF 62  
 Db 7 RRSVLVFPFAQGHISPMQMLAKTLHLKGFSTIVVQTKF-----NYFSPSDQTFDHFQF 60

QY 63 VAFSDGYDD-----GLKPCGDKRYMSEMKARGSEALRNLL-NHHVDYTFVYVSHLFAWAA 117  
 Db 61 VTPESLPESDFKMLGPIQLFKLNKECKVSFKDCILGQLVLQSGNEISCVIYDFEYFAE 120

QY 118 EVARESOPVALLWPEPATVLCIYFYFYNGYADEIDA-----GSDEIOLPRLPPLPEQRS 171  
 Db 121 AAARKEKLPNIISTTSATAFACRSVDFDKLIYANNVOAPLKETGQOEELVPEFPPLRYKD 180

QY 172 LPTFLLPETPERF---RLMMKEKLETLGDEEKAKVLVNTFDALPDALTAIDRYEL---- 224  
 Db 181 FPV-----SREASLESIMEVYRNTVKRTASSVIINTASCLSSLSLFLQOOLQIPV 233

QY 225 IGTGPL-----IPSAFLDGGDPSETS YGGDLFEKSEENNCVWLDTPKSSVYVVSFGSV 279  
 Db 234 YPIGPHLMVASPTSLLE-----ENKSCIEWLNKQKVNVIYISMGSI 276

QY 280 LRFPKQAMEIGIGLACGRPFLMIREQKNDGDEEEELSCIGELKMK-----GKTVSW 335  
 Db 277 ALMEINEIMEVASGLAASNOHFLWIRPGSIPGSEWIE---SMPEFSKMLVDGRIYVKW 333

QY 336 CSOLEVLAHPALGCFVTHCGNVSATESUSCGVPVAVPQWFDQTTNAKLIEDAWGTQVVR 395  
 Db 334 APQKEVLHPAVGFGVSHCGWNTLESIGQVPMICRPFSGDKQVNAVRYLECVMKIGIQV 393

QY 396 RMNEGGVDGSETERCVE--MVMDGGEKSLVRENAIKWKTLEAREAMGEGSSLKLNALF 453  
 Db 394 E-----GELDRGVVERAVKRLMVDDEGEE---MKRRAPSLREQLRASVKGSGSHNSLEEF 446

QY 454 LH 455  
 Db 447 VH 448

RESULT 13  
 T00584  
 indole-3-acetate beta-glucosyltransferase homolog T27E13.12 - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999  
 C;Accession: T00584  
 R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K  
 submitted to the EMBL Data Library, May 1998  
 A;Description: Arabidopsis thaliana chromosome II BAC T27E13 genomic sequence.  
 A;Reference number: Z14177  
 A;Accession: T00584  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-455 <ROU>  
 A;Cross-references: EMBL:AC004165; NID:g3150396; PID:g3150406  
 A;Experimental source: cultivar Columbia  
 C;Genetics:  
 A;Map position: 2  
 A;Introns: 166/1  
 A;Note: T27E13.12  
 C;Superfamily: flavonol O3-glucosyltransferase

Query Match 20.0%; Score 484; DB 2; Length 455;  
 Best Local Similarity 30.6%; Pred. No. 2.7e-31;  
 Matches 147; Conservative 71; Mismatches 196; Indels 66; Gaps 17;

QY 4 RRVLLATPFAQGHINPALQFAKRLK--ACTDVTFTTSVYAWRRMANTASAAAGNPPGLD 61  
 Db 12 RHVVAMPYPCRGHINPMNLCRLVRRYPNLHVTFTVVT-EEW-----LGFTGDPDKPD 63

QY 62 FVAFS-----DGYDDGLPCGDKRYMSEMKARGSEALRNLL--LNHHVDYTFVYVSHLFAWA 116  
 Db 64 RIHFSTLPNLIPSELVRAKDFIGFIDAVYTRLEPPEKLLDLSNPPSVIFADTVYVIA 123

QY 117 AEVARESOPVALLWPEPATVLCIY-----YFYNGYADEIDAGSDEI--QLPRLPPELQ 170  
 Db 124 VRGRKNIPVSWLWNTMSATILSFFLHSDLLISHGHA-LFEPSEEVVDVVPGLSPTKLR 182

QY 171 SLPTFLLPETPERFLMMKEKLETLGDEEKAKVLVNTFDALP-----DALTAIDRYELIGI 227  
 Db 183 DLPP-IPDGYSDRVFTKLCFDELPGAR--SLIFTAYELEHKAIDAFSKLDIPYAI 239

QY 228 GPLPSAFLDGGDPSETS YGGDLFEKSEENNCVWLDTPKSSVYVVSFGSVLRFPPKAQM 287

Db 240 GPLIPF-----EELSVQND-----NKEPNYIQWLEQPEGSLVYISQGSFLSVSEAQM 287  
Qy 288 EEIKGGLLACGRPLWMTREOKNDGEEBEELSCIGBLK-----KMGKIVSWCSOLE 340  
Db 288 EEIVKGLRESGVRLVARG-----GSLKLEALEGSLGVVWVWCDQLR 331  
Qy 341 VLAHPALGCFVTHCGWNSAVESLSCGVPVAVPOWFQDQTTNAKLIEDAWGTGVRVRNNEG 400  
Db 332 VLCHKAVGGFWTHCGFNSTLGGIYGVPLAFPLFWDQILNAKMLVEDWRVGMRIERTKK 391  
Qy 401 GG--VDGSEIERCEVEMVMD--GGEKSKLVRENAIKWTKTLAREAMGSDGSKLNLAFLH 457  
Db 392 NELLIGREEIKVKKRFNDRESEGEKEMRRACDLSEISRGAVAKSGSSNVNIDEFVRHI 451  
RESULT 14  
T45604  
glucosyltransferase-like protein - Arabidopsis thaliana  
N:Alternate names: protein F12A12.190  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000  
C:Accession: T45604  
R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; submitted to the Protein Sequence Database, December 1999  
A:Reference number: 223008  
A:Accession: T45604  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-451 <CHO>  
A:Cross-references: EMBL:AL13314  
A:Experimental source: cultivar Columbia; BAC clone F12A12  
C:Genetics:  
A:Map position: 3  
A:Introns: 160/1  
A:Note: F12A12.190  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 19.3%; Score 466.5; DB 2; Length 451;  
Best Local Similarity 29.5%; Pred. No. 6.9e-30;  
Matches 140; Conservative 72; Mismatches 205; Indels 57; Gaps 13;  
Qy 4 RRVLLATFPAGGHINPALQFAKRLKAGTDVTFSTSVYAWRMANTASAAAGNPGLDFV 63  
Db 8 RRVLVAVPAGGHISPIHQLAKTLHLKGFSTIATQKFN-----FSPSDDFT 55  
Qy 64 AF-----SDGYDDGLKPCGDKRYMSEMKARGSEALRNLL--NNHDVTFVYVYGH 112  
Db 56 DFQVFTIPESLPESDFED--LGPIELHLKNECVSFKDCGLQLLQOQNEIACVYDEF 114  
Qy 113 FAWAAEAVARESQVPSALLWVEPATVLCIYYFYFNGYADEI-----DAGSDEIQPLRPP 166  
Db 115 MYFAAAAKEFKLPNVIFSTTSATAFVCRSAFDKLYANSILTPLEKPGQQONELVPEFHP 174  
Qy 167 LEQSLPTFLPETPERFLMKKEKLETLDGEEKAKVLVNTFDALPTAIDRYELIG 226  
Db 175 LCKDFPV-----SHWASLESMMELYRNTVDKRTASSVIINTASCLESLSLKLQQLQIP 230  
Qy 227 IGPLTSAFLDGGDPSETSYGDLFEKSENNCVLWDTKPKSSVYVYVYFSGSVLFRPPK 286  
Db 231 VYPIGPHLVASASTS-----LLE--ENKSCIEMLNKKQKNSVIFVSLGSLALMEINE 281  
Qy 287 MEETKGLLACGRPLWMTREOKNDGEEBEELSCIGBLK-----KMGKIVSWCSOLEVLA 343  
Db 282 VIETALGLDSSKQQLVIRPGSVRGSEWLENLPREFSKI--ISGRYIVKMAPOKEVLS 339  
Qy 344 HPALGCFVTHCGWNSAVESLSCGVPVAVPQWFQDQTTNAKLIEDAWGTGVRVRNNEG 403  
Db 340 HPVAGGFWSHCGWNSLTLESIGEGVPMICKPFSSDQMVNARYLECVWKIGIQVE----GD 395  
Qy 404 DGETERCE--MVMDGGEKSKLVRENAIKWTKTLAREAMGSDGSKLNLAFLH 455

Db 396 DRGAVERAVRRLMWEEEGEG---MKRAISLKEQLRASVISGGSSHNSLEEFVH 446  
RESULT 15  
T12981  
hypothetical protein T6H20.280 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 29-Oct-1999  
C:Accession: T12981  
R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; submitted to the Protein Sequence Database, July 1999  
A:Reference number: Z17586  
A:Accession: T12981  
A:Molecule type: DNA  
A:Residues: 1-452 <CHO>  
A:Cross-references: EMBL:AL096859; GSPDB:GN00061; ATSP:T6H20.280  
A:Experimental source: cultivar Columbia; BAC clone T6H20  
C:Genetics:  
A:Gene: ATSP:T6H20.280  
A:Map position: 3  
A:Introns: 161/1  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 18.9%; Score 458.5; DB 2; Length 452;  
Best Local Similarity 29.2%; Pred. No. 3.1e-29;  
Matches 141; Conservative 80; Mismatches 193; Indels 69; Gaps 16;  
Qy 3 RRVLLATFPAGGHINPALQFAKRLKAGTDVTFSTSVYAWRMANTASAAAGNPGLDF 62  
Db 7 KRRIVLPVAAQGHVTPMQLGKALQSKGFLIT-----VAQRQFNQIGSSLSQHPGDF 60  
Qy 63 VAFSDGYDDGLKPCGDKRYMSEMKARG--SEALRNL-----LLNNHDVTF 105  
Db 61 VTIPESLP-----QSESKKLGPAEYLMNLNKTSEASFKECISQLSMQGGNDIA 108  
Qy 106 FVYVSHLFAWAAEAVARESQVPSALLWVEPATVLCIYYFYFNGYADE--IDAGSDEIQ--- 160  
Db 109 CIIYDKLMYFCEAAAKEFKIPSVIFSTSATIQVCYCVLSAEKFLIDMKDPEKQKV 168  
Qy 161 LRLPPLPQRSLPFT--LLPETPERFLMKKEKLETLDGEEKAKVLVNTFDALPDALTAI 219  
Db 169 LEGLHPLRYKDLPTSGFGFLEP-----LLEMCREVVNKRTASAVIINTASCLESLSWL 223  
Qy 220 DRYELIGIPLIPSAFLDGGDPSETSYGDLFEKSENNCVLWDTKPKSSVYVYFSGV 279  
Db 224 QOELGIPYPLGLPHI-----TASSPGPSLIQ--EDMSCIEMLNKKPRSVIYISLGTK 275  
Qy 280 LRFRAQMEETGKGLLACGRPLWMTREOKNDGEE---EELSCIGELKMGKIVSWC 336  
Db 276 AHMETKEMLEHAWGLNLSNQPLWIRPGSVAGFWIELLPEV--IKMWTERTGIYAKWA 333  
Qy 337 SOLEVLHPALGCFVTHCGWNSAVESLSCGVPVAVPQWFQDQTTNAKLIEDAWGTGVRV 396  
Db 334 PQIEVLGHPAVGFWSHCGWNSLTLESIGEGVPMICKPFSSDQMVNARYLECVWKIGIQ 393  
Qy 397 MNEGGVVGDSETERCE--MVMDGGEKSKLVRENAIKWTKTLAREAMGSDGSS---LKNLNA 452  
Db 394 ----GEVEREGYERAKRLIID--EGAAMRERALDLKEKLNASVRSGSSYNALDELVK 447  
Qy 453 FLH 455  
Db 448 FLN 450

Search completed: August 1, 2000, 21:27:44  
Job time: 11212 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 13:52:00 ; Search time 47.63 Seconds  
(without alignments)  
228.755 Million cell updates/sec

Title: US-09-147-955-2  
Perfect score: 2421  
Sequence: 1 MYRRVLLATFPAQGHINPA.....GEDGSLKLNALHQAVERA 460

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2421	100.0	460	1 W92947	W09905287 Seq ID 7
2	2223	91.8	443	1 W92948	W09905287 Seq ID 8
3	1646.5	68.0	461	1 W92949	W09905287 Seq ID 9
4	1445	59.7	478	1 W92950	W09905287 Seq ID 1
5	1363.5	56.3	468	1 W92952	W09905287 Seq ID 1
6	762.5	31.5	471	1 W03736	Maize UDP-glucose
7	700	28.9	466	1 W09843	UDP-glucose:thiohy
8	407.5	16.8	471	1 W98009	Medium-chain UDP-g
9	406.5	16.8	470	1 W47172	Glucosyl transfera
10	398	16.4	433	1 W13406	Solanum melongena
11	391	16.2	488	1 W64569	Potato solanidine
12	369.5	15.3	473	1 W56431	UDP-glucose:flavon
13	326.5	13.5	471	1 R49245	Anthocyanidin-3-gl
14	145.5	6.0	534	1 R26154	HUG-Br2. Isolated
15	139.5	5.8	533	1 R26153	HUG-Br1. Isolated
16	121.5	5.0	506	1 R10429	Ecdysteroid UDP-gl
17	121.5	5.0	530	1 W47126	Uridine diphospho-
18	119	4.9	515	1 W56750	Ecdysteroid UDP-gl
19	109	4.5	414	1 R93982	Saccharothrix aro
20	107	4.4	74	1 R30165	UGT1 Exon 4 produc
21	107	4.4	421	1 W99388	S. erythraea desosa
22	105.5	4.4	409	1 W71464	Glycosyltransferas
23	101	4.2	473	1 W64392	S. cerevisiae 1947
24	101	4.2	1198	1 W64384	S. cerevisiae 1947
25	100	4.1	421	1 W19735	Sugar biosynthesis
26	98	4.0	399	1 R13989	Zeaxanthin glycosy
27	94	3.9	508	1 W35712	Torenia flavonoid
28	93	3.8	580	1 W90217	Bispecific tetra
29	88	3.6	431	1 Y00181	Enterococcus faeca
30	88	3.6	474	1 Y00180	Enterococcus faeca
31	87.5	3.6	781	1 W11298	Phosphofructokinase
32	87.5	3.6	1051	1 W52304	Glucodextranase pr
33	87	3.6	426	1 W99397	S. antibioticus des

ALIGNMENTS

RESULT 1

W92947  
ID W92947 standard; Protein; 460 AA.  
AC W92947; 1999 (first entry)  
DT 14-MAY-1999 (first entry)  
DE W09905287 Seq ID 7.  
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.  
OS Perilla frutescens.  
PN W09905287-A1.  
PD 04-FEB-1999.  
PF 16-JUL-1998; J03199.  
PR 25-JUL-1997; JP-200571.  
PA (SUNR) SONTORY LTD.  
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
PI Yamazaki M;  
DR WPI: 99-142940/12.  
DR N-PSDB; X02826.  
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
PT used to transform plants for improvement of plant coloration  
PS Claim 2; page 71-73; 89pp; Japanese.  
CC This invention describes the isolation of plant proteins which have  
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used  
CC in the production of plant varieties with improved colour, for example in  
CC the production of flower varieties with desired coloration and cut  
CC flowers harvested from them.  
SQ Sequence 460 AA;

Query Match 100.0%; Score 2421; DB 1; Length 460;  
Best Local Similarity 100.0%; Pred. No. 4.8e-248;  
Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYRRVLLATFPAQGHINPALQFAKRLKAGTDVTFFTSVYAWRRMANTASAAAGNPGL 60  
Db 1 MYRRVLLATFPAQGHINPALQFAKRLKAGTDVTFFTSVYAWRRMANTASAAAGNPGL 60  
Qy 61 DFVAFSDGYDGLKPCGDKRYMSEMKARGSEALNLLNNHNDVTFVYVYSHLFAAAEYA 120  
Db 61 DFVAFSDGYDGLKPCGDKRYMSEMKARGSEALNLLNNHNDVTFVYVYSHLFAAAEYA 120  
Qy 121 RESQVPSALLWVEPATVLCIYFFYNGYADEIDAGSDEIQPLRPLPQRSPLTFLPET 180  
Db 121 RESQVPSALLWVEPATVLCIYFFYNGYADEIDAGSDEIQPLRPLPQRSPLTFLPET 180  
Qy 181 PERFLAMKEKLETLDGSEKAKVLVNTFDALPDALTADRYELIGIGLPLPSAFLDGSD 240  
Db 181 PERFLAMKEKLETLDGSEKAKVLVNTFDALPDALTADRYELIGIGLPLPSAFLDGSD 240  
Qy 241 PSETSYGGDLFEKSENNCVWLDTPKSSVYVYVSGVLPFPKQMEIEIGKLLACGRP 300  
Db 241 PSETSYGGDLFEKSENNCVWLDTPKSSVYVYVSGVLPFPKQMEIEIGKLLACGRP 300  
Qy 301 FLWMIREQNDDGEDEEELSCIGELKMKGIKIVSWCSQLEVLHAHPALGCFVTHCGWNSAV 360  
Db 301 FLWMIREQNDDGEDEEELSCIGELKMKGIKIVSWCSQLEVLHAHPALGCFVTHCGWNSAV 360

A. thaliana SGT pr  
A. thaliana SGT pr  
A. thaliana protel  
Mesotaenium caldar  
Grapevine leafroll  
Polypeptide with e  
Protein encoded by  
ICAM-R (intercellu  
S. tuberosum isoam  
Human type 2 neuro  
Mutant thermostabl  
Mutant thermostabl

```

QY 361 ESLSGCVPPVAVPQWFQDQTTNAKLIEDAGTGVVRMNEGGVDSGEIERCVEMVMDGGE 420
Db 361 ESLSGCVPPVAVPQWFQDQTTNAKLIEDAGTGVVRMNEGGVDSGEIERCVEMVMDGGE 420
QY 421 KSKLVRENAIKWKTAREAMGEGDSSKLNLNAPLHQVARA 460
Db 421 KSKLVRENAIKWKTAREAMGEGDSSKLNLNAPLHQVARA 460

RESULT 2
W92948
ID W92948 standard; Protein: 443 AA.
AC W92948;
DT 14-MAY-1999 (first entry)
DE W0905287 Seq ID 8.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
OS Ferlilia frutescens.
PN W0905287-Al.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR ) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
PI WPI; 99-142940/12.
DR N-PSDB; X02827.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PT used to transform plants for improvement of plant coloration
PS Claim 2; Page 74-76; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC the production of flower varieties with desired coloration and cut
CC flowers harvested from them.
SQ Sequence 443 AA;

Query Match 91.8%; Score 2223; DB 1; Length 443;
Best Local Similarity 95.7%; Pred. No. 4.4e-227;
Matches 424; Conservative 9; Mismatches 8; Indels 2; Gaps 1;

QY 1 MVRRLVLLATFPAGQHINPALQFAKRLKAGTDVTFSTVYAWRRMANTASAAAGNPPGL 60
Db 1 MVRRLVLLATFPAGQHINPALQFAKRLKAGTDVTFSTVYAWRRMANTASAAAGNPPGL 60
QY 61 DFVAFSDGYDDGLKPGDGKRYMSEMKARGSEALRNLLNNHDDVTFVYVSHLFA 120
Db 61 DFVAFSDGYDDGLKPGDGKRYMSEMKARGSEALRNLLNNHDDVTFVYVSHLFA 120
QY 121 RESQVPSALLWVEPATVLCIYFFNGYADEIDAGSDEIOLPRLPLEQSLPTFLLPET 180
Db 121 RLSHVPALLWVEPATVLCIYFFNGYADEIDAGSDEIOLPRLPLEQSLPTFLLPET 180
QY 181 PERFLRMKKEKLETLGEEKAKVLTNTFDALPDALTAIDRYELIGIPLIPSAFLDGGD 240
Db 181 PERFLRMKKEKLETLGEEKAKVLTNTFDALPDALTAIDRYELIGIPLIPSAFLDGGD 240
QY 241 PSTSTYGGDLFEKSEENNCVEMLDTPKSSVYVSGSVLRFPPKAQMEETIGKGLACGRP 300
Db 241 PSTSTYGGDLFEKSEENNCVEMLDTPKSSVYVSGSVLRFPPKAQMEETIGKGLACGRP 300
QY 301 FLWMIREQKNDDG--EPEEELSCIGELKKMGKIVSMCSQLEVLHAHPALGCFVTHCGWNS 358
Db 301 FLWMIREQKNDDGEEEEEELSCIGELKKMGKIVSMCSQLEVLHAHPALGCFVTHCGWNS 360
QY 359 AVESLSGCVPPVAVPQWFQDQTTNAKLIEDAGTGVVRMNEGGVDSGEIERCVEMVMDG 418
Db 361 AVESLSGCVPPVAVPQWFQDQTTNAKLIEDAGTGVVRMNEGGVDSGEIERCVEMVMDG 420
QY 419 GEKSKLVRENAIKWKTAREAMG 441
Db 421 GDKTKLVRENAIKWKTAREAMG 443

```

```

RESULT 3
W92949
ID W92949 standard; Protein: 461 AA.
AC W92949;
DT 14-MAY-1999 (first entry)
DE W0905287 Seq ID 9.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
OS Verbena hybrida.
PN W0905287-Al.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR ) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
PI WPI; 99-142940/12.
DR N-PSDB; X02828.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PT used to transform plants for improvement of plant coloration
PS Claim 2; Page 76-78; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC the production of flower varieties with desired coloration and cut
CC flowers harvested from them.
SQ Sequence 461 AA;

Query Match 68.0%; Score 1646.5; DB 1; Length 461;
Best Local Similarity 70.0%; Pred. No. 6.2e-166;
Matches 326; Conservative 49; Mismatches 74; Indels 17; Gaps 7;

QY 1 MVRRLVLLATFPAGQHINPALQFAKRLKAGTDVTFSTVYAWRRMANTASAAAGNPPGL 60
Db 1 MSRAHLLATFPAGQHINPALQFAKRLANADIQVTFSTVYAWRRMANTASAAAGNGLI 57
QY 61 DFVAFSDGYDDGLKPGDGKRYMSEMKARGSEALRN--LLNNHDDVTFVYVSHLFA 114
Db 58 NFVFSFGYDDGLQPGDDGKRYMSEMKARGSEALRN--LLNNHDDVTFVYVSHLFA 117
QY 115 WAAEVAAREQVPSALLWVEPATVLCIYFFNGYADEIDAGSDEIOLP-RLPPLERSLP 173
Db 118 WAAKVAAREFRLSALLWIEPATVLDIFVFFNGYDEIDAGSDAIHLPGSLPVLQARDLP 177
QY 174 TFLLPETPERFLRMKKEKLETLGEEKAKVLTNTFDALPDALTAIDRYELIGIPLIPS 233
Db 178 SFTLPSTHERFRSLMKEKLETLGEEKAKVLTNTFDALPDALTAIDRYELIGIPLIPS 237
QY 234 AFLDGGDPSSTSYGGDLFEK--SEENNCVEMLDTPKSSVYVSGSVLRFPPKAQMEETIG 292
Db 238 AFLDGGDPSSTSYGGDLFEKSGNDLCLWLTSTNPRSSVYVSGSVLRFPPKAQMEETIG 297
QY 293 GLACGRPFLLWMIREQKNDDGEEEEEELSCIGELKKMGKIVSMCSQLEVLHAHPALGCFVT 352
Db 298 GLDCCRPFLWVVRV-----NEGEEVLISCEELKRVGKIVSMCSQLEVLTHPSLACGFT 352
QY 353 HCGWNSAVESLSGCVPPVAVPQWFQDQTTNAKLIEDAGTGVVRMNEGGG-VDSSETERC 411
Db 353 HCGWNSVLESISGVPVAVPQWFQDQTTNAKLIEDAGTGVVRMNEGGG-VDSSETERC 412
QY 412 VEMVMDGGEKSKLVRENAIKWKTAREAMGEGDSSKLNLNAPLHQV 457
Db 413 IEEVMDGGEKSKLVRENAIKWKTAREAMGEGDSSKLNLNAPLHQV 458

RESULT 4
W92950
ID W92950 standard; Protein: 478 AA.
AC W92950;
DT 14-MAY-1999 (first entry)
DE W0905287 Seq ID 10.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.

```







salicylic acid; jasmonic acid; ethylene; wound inducible gene;  
plant defence protein; plant response; tobacco; rice.  
Lycopersicon sp.  
WO9745546-A1.  
PN  
04-DEC-1997.  
PD  
30-MAY-1997; G01473.  
PF  
31-MAY-1996; GB-011420.  
PR  
(UYYO-) UNIV YORK.  
PA  
Bowles DJ, Calvert CM, ODonnell PJ, Roberts MR;  
WPI; 98-032653/03.  
DR  
N-PSDB; V17054.  
DR  
Tomato wound inducible (TWI1) gene encoding glucosyl transferase -  
PT  
useful to develop products that alter signalling pathways in plants  
PT  
by altering of salicylic acid, jasmonic acid or ethylene  
PS  
Claim 2; Fig 3; 52pp; English.  
CC  
This is a glucosyl transferase (Grase) protein encoded by a wound  
CC  
inducible gene (TWI1) isolated from wounded tomatoes. The TWI1 gene  
CC  
encodes this Grase from amino acid position 5. The TWI1 gene can be  
CC  
used to identify homologue Grase encoding genes isolated from tobacco  
CC  
and rice. A microbial host can be transfected or transformed with a  
CC  
vector containing the Grase encoding nucleic acids. The products can be  
CC  
used to interfere with Grase and therefore alter signalling pathways in  
CC  
plants, specifically tobacco, rice or tomato plants by altering levels  
CC  
of salicylic acid, jasmonic acid or ethylene. This can induce the  
CC  
production of plant defence proteins such as pathogenesis-related (PR)  
CC  
and proteinase inhibitor (PIN) proteins which regulate plant development  
CC  
(plant growth, reproduction and senescence) and improve plant response to  
CC  
pathogens.  
SQ  
Sequence 470 AA;

```

Query Match      16.98; Score 406.5; DB 1; Length 470;
Best Local Similarity 28.3%; Pred. No. 1.9e-34;
Matches 142; Conservative 64; Mismatches 176; Indels 119; Gaps 21;

QY      11  FP--AOGHINPALQFAKRLKAGTDVTF--SYAWRRMAN-----TAS 51
Db      9  FPDAOGHMIPTLDMANVACRGVKATITTPLNESVFSKAIERKHLGIEIDIRLLKFP 68
QY     52  AAAGNPP---GLDFVAFSDGYDDGLKPCGDKGRYKSEMARGSEALRNLLLNHHVDVF 107
Db     69  AKENDLPEDCERLDLVPSDDKLPNFLAKAAMKDFEE-----LIGCRPPDL 116
QY     108  VYSHLFMAAEVARESQVPALLWEPATVLCIYYFYFNGYADEIDAGSDIEQIPRLPL 167
Db     117  VSDMFLPWTYDAAKFSIPRIVFHGTSYFALCV-----GDTIR--RNKPF 159
QY     168  EQRS--LPTLLPPTPERFLMKKE--KLEULD-----GREKAK--VLVNTF 208
Db     160  KNVSSDTETFPVDPDLPHEIRLTRTQLSPFEQSDDEETGMAPIKAVRESDAKSYGVIFNSF 219
QY     209  DALEPDALTADRY-ELIG-----IGPLIPSAFLDGGDPSETSYGGDLFEKSE----- 255
Db     220  YELESD--VYEHYTKVGRKNAIGPL-----SLCNRDIEDKARCKSSI 263
QY     256  -ENNCVENLOTKPKSSVYVYSGSVLRAPPKAQMBEIKGLLACGRPFLMMIREQND--- 311
Db     264  DEHACKLWDSKKSSIVYVYCGFTADFTTAQMOELAMGLASGQDFIWTIRTGNEDWLP 323
QY     312  DGEEREEELSCIGELKWKGTKIV-SWCSOLEVLNHLAPGCGFTHCGWNSAVESLSGVPDV 370
Db     324  EGPEE-----RTKEKGLIIRWAPQSVILDHIEAGAFVTHCGWNSLTLEGISAGVPMV 375
QY     371  AVPQWFQDTTNAKLIEDAWGTGVRV-----RNNSEGGYDGSIEIERCVMWMDGGEKSLV 425
Db     376  TWPVFAQGFNEKLVTEVMSRGAGVSGKQWKRTASEGVKRAIAKAIKRVM-ASEETEGF 434
QY     426  RENAIKWKTILAREAMGEDGSS 446
Db     435  RSRakeYKEMAREAIERGGSS 455

```

RESULT 10

W13406  
ID W13406 standard; Protein; 433 AA.  
AC W13406;  
DT 10-JUN-1997 (first entry)  
DE Solanum melongena flavonoid-3-glucosyl transferase.  
KW Eggplant; flavanoid-3-glucosyl; transferase; glucose; flavonoid;  
KW pigment; production; seed; plant.  
OS Solanum melongena.  
PN J09056385-A.  
PD 04-MAR-1997.  
PF 25-AUG-1995; 238943.  
PR 25-AUG-1995; JP-238943.  
PA (KIRI ) KIRIN BREWERY KK.  
DR MPI; 97-208631/19.  
DR N-PSDB; T63123.  
PT Solanum melongena flavonoid-3-glucosyl transferase and DNA encoding  
PT it - catalyses transfer of glucose to 3-position of flavonoid  
PT pigment in plants  
PS Claim 4; Pages 9-10; 14pp; Japanese.  
CC The present sequence is the Solanum melongena (eggplant)  
CC flavanoid-3-glucosyl transferase [P3GT], which catalyses the  
CC transfer of glucose to the 3-position of flavonoid pigments. The  
CC F3GT DNA is useful for the production of the seed of a plant. A  
CC S. melongena seed was cultured under a red light and a flavanoid  
CC derived. A cDNA library in which the pigment was concentrated was  
CC prepared. A flavanoid glucosidising enzyme gene was cloned and its  
CC structure and mRNA expression pattern analysed. The clone was  
CC expressed in E. coli.  
SQ Sequence 433 AA;

Query Match	16.48;	Score 398;	DB 1;	Length 433;
Best Local Similarity	26.7%;	Pred. No. 1.3e-33;		
Matches 124;	Conservative 87;	Mismatches 176;	Indels 78;	Gaps
QY	6	VLIATPQAQHIIPALQFAKRL---	LKAGTDVTFSTSVAMRMANTASAAAGNPPGLDP	62
	:			:
	:			:
Db	8	IAELAFPGFTHATPLTLVQKISPELPSSTIFSFNT---	SSNSISFSPVQEN	60
	:			:
	:			:
QY	63	VAFSDGVDDLKPCGGDKRYMSEKARGSEALR---	NLLNHN---HDVTFV---	108
	:			:
Db	61	IKIYNW-DGVKGGND-----	TPFGEAIKLFQSTLLSKITTEABEETGVKFS	111
	:			:
QY	109	YSHLFAWA--AEVARESQVP	SALLVEPAPVLCIYFYFYNGYADEIDAGSD-EIQLPRL-	164
	:			:
Db	112	FSDAFLWCLVLKPKMNAPGVAYWTGGSCSLAVHL-----	YDILRSNKETSLKIPGFS	166
	:			:
QY	165	PPLEQRSPLFTLPE-TPERFRLMMEKL--ETLOGEEKAKVLVNTFD	DALEPDALTAID-	220
	:			:
Db	167	STLSINDIP---PEVTAEDLEGPMSSMLYNMALNLHKA	DAVVLNSFQELDRPOLINKDL	222
	:			:
QY	221	---RYELIGIGLPIPSAFLDGGDPSTSYGGDILPEKSEENNC	VEWLDTRPKSSVYVYSG	277
	:			:
Db	223	QKNLQKVFNTGPIVLQVS-----	SRKLDGSGCQWLDDKQEKSVVYLSEF	266
	:			:
QY	278	SVLRFPKQAEIEGKGLLACACRPF	LMWIREQKNDGDEEEELSCIGELKMGKIVSWCS	337
	:			:
Db	267	TVTTLPPNEIGSTAEALETKTKTPFIWSLR---	NNGVKNLPK-GPIERTKEFGKIYSWAP	321
	:			:
QY	338	QLEYLAHPALPGCFVTHCGWNSAVYESLS	CGVPVAVPQWFQDOTTNAKLIEDAWGTGVRVM	397
	:			:
Db	322	QLEILAHKSVGVFVTHCGWNSILEGISFGVPMICRPF	FGDQKLNSRWVESVWEIGLQI--	379
	:			:
QY	398	NEGGVDGSIERCVMVMDGGGKSLVRENAIKWKT	TLAREANGE	442
	:			:
Db	380	-EGIFTKSGIISALDITFFN-EKGGILIRNVEGLKEA	LEAVNO	422
	:			:

RESULT	11
W64569	
ID	W64569
AC	W64569
DT	07-DEC

DE Potato solanidine UDP-glucose glucosyltransferase.  
KW Solanidine UDP-glucose glucosyltransferase; SGT; potato; tomato;  
glycoalkaloid; solasodine; solanaceous plant; transgenic plant.  
OS Solanum tuberosum cv. Lemhi Russet.  
FH Location/Qualifiers  
FT Region 12..46  
FT /note="conserved domain"  
FT Region 110..143  
FT /note="putative steroid binding region"  
FT Region 351..401  
FT /note="UDP-glucose binding region"  
PN W09834471-A1.  
PD 13-AUG-1998.  
PR 30-JAN-1998; U01864.  
PR 07-FEB-1997; US-797226.  
PA (USDA ) US SEC OF AGRIC.  
PI Allen PV, Belknap W, Friedman M, Moehs CP, Rockhold DR,  
PI Stapleton A;  
PI WPI; 98-446828/38.  
DR N-PSDB; V49609.  
PT New DNA encoding solanidine UDP-glucose glucosyl-transferase and  
PT related vectors - and transgenic plants, used to reduce contents of  
PT steroidal glyco-alkaloid(s) in plants, specifically potato  
PS Example 1; Fig 2; 54pp; English.  
CC This is the amino acid sequence of potato cv. Lemhi Russett  
CC solanidine UDP-glucose glucosyltransferase (SGT). It was deduced  
CC from an SGT cDNA sequence (see V49609). SGT is involved in the  
CC biosynthesis of steroidal glycoalkaloids in solanaceous plants,  
CC catalysing the UDP-glucose dependent conversion of the aglycone  
CC solanidine to gamma-chaconine. The invention relates to DNA  
CC sequences which encode SGT, and their use, particularly use of an  
CC antisense DNA construct to inhibit SGT activity and glycoalkaloid  
CC levels in solanaceous plants. Transgenic plants are claimed,  
CC particularly tomato and potato, that have reduced contents of toxic  
CC steroidal glycoalkaloids. The inactivation of glycoalkaloid  
CC biosynthetic pathways is beneficial to reduce or eliminate  
CC glycoalkaloid biosynthesis during storage and shipping.  
SQ Sequence 488 AA;

Query Match 16.2%; Score 391; DB 1; Length 488;  
Best Local Similarity 23.9%; Pred. No. 8.7e-33;  
Matches 125; Conservative 89; Mismatches 186; Indels 122; Gaps 20;  
QY 6 VLLATPAGHINPALQFAKRLKAGTDVTFVTSVYAWRRMANT----- 49  
DB 13 VLFLPFSAGHIFPLVNAARFASRGVAKTILTPHNALLFRSTIDDDVRISGPPISIVT 72  
QY 50 ---ASAAAGNPPGLDFVAFSDGVDGLKPGCGDKRKY--MSEMKARGSEALRNLLNNHDV 104  
DB 73 IKPFAEVLGPEGI-----ESFNSATSPENPHKIFVALSLQKPMEDXIRELRPD--- 122  
QY 105 TFVYVYSHL-FAMAAEAVARESQVPSALLWVEPATVLCIYYFY-FNGYADEIDAGSDTQLP 162  
DB 123 --CIFSDMYEPWTVDIADELHIPRL--YNLSAYMCYSIMHNLKVVYRPHKQPNLDESQ-- 176  
QY 163 RLPLPLEORSPLTLLPTPERFLMMKEKLETL-----DGEKAKVL 204  
DB 177 -----SFVPGPLPEIKFKLSQLTDLRKSDQDKTVFDELLEQVEDSEERSYGI 225  
QY 205 V-NTFDALPDALTAIDRYELI-----GIGLIPSAFLDGGDPSSETSYGDLFEK---S 254  
DB 226 VHDTFYELEP---AYVDYYOKLKKPKCWHFGL-----SHPASKINSKELIS 269  
QY 255 EENN---CYEWLDTKPKSSVYVYSGVLRFPPKAQMEIEIGKLLACGRPFLMWIREQKND 311  
DB 270 EHNNEIIVDLWNAQPKSVLYVSGFSGMARFPESQLEIAQALDASNVPTFVLRP--- 325  
QY 312 DGEEEELSCIGELKMKXK-----IVSWCSQLEVLAIHPALGCFVTHCGWNSAVESLSGCV 367  
DB 326 --NEETASWLPVGNLEDKTKGLYIKGVVPLQITIMEHSATGGFTGCTNSVLEAITFG 383  
QY 368 PVVAVPQWFDQTTNAKLIE-----DAWGTGVVRVRNREGGVDGSEIERCIVE--MV 415

DB 384 PMITWPLVADQFYNEKVVYVRGLGKIGKIDVWNEGIEI---TGPVIESAKIERALRLMI 440  
QY 416 MDGGEKSKLVRENAIKWKTTLAREAMGEDGSLKLNLAFLHQV 457  
DB 441 SNGSEETIINDRYMAKMAQNAATNEGSSWNLTALIOHI 482  
RESULT 12  
W56451  
ID W56451 standard; Protein; 473 AA.  
AC W56451;  
DT 06-AUG-1998 (first entry)  
DE UDP-glucose:flavonoid 3,5-O-glucosyl transferase amino acid sequence.  
KW UDP-glucose:flavonoid 3,5-O-glucosyl transferase; gentian; rose plant;  
KW blue flower.  
OS Unidentified.  
PN J10113184-A.  
PR 15-OCT-1996; 272253.  
PR 15-OCT-1996; JP-272253.  
PA (TOLG ) TOKYO GAS CO LTD.  
DR WPI; 98-315473/28.  
DR N-PSDB; V23108.  
DR Gentian flavonoid glucosyl transferase gene - used to produce  
PT recombinant rose plants with blue petal(s)  
PS Claim 1; Pages 6-7; 10pp; Japanese.  
CC The present sequence represents a UDP-glucose:flavonoid 3,5-O-glucosyl  
CC transferase enzyme. The nucleic acid sequence was amplified from  
CC RNA which had been isolated from gentian (sic) petals. The gene is  
CC used to transform rose plants to produce blue flowers.  
SQ Sequence 473 AA;  
Query Match 15.3%; Score 369.5; DB 1; Length 473;  
Best Local Similarity 25.7%; Pred. No. 1.6e-30;  
Matches 131; Conservative 90; Mismatches 199; Indels 89; Gaps 18;  
QY 1 MYRRVLLATPFAQGHINPALQFAKRLKAGTDVTF-----TSVYAWRRMANTAS 51  
DB 1 MKKAELVIIPFGISHLGSTVELAKLAEARNEHLSIVIIKFPNDTKVSNLLSLSTAS 60  
QY 52 AAAGNPPGLDFVAFSDGVDGLKPGCGDKRYMSEMKARGSEALRNLLNNHDVTFVYVS 110  
DB 61 RIKVIELKQETISYETG-----PLFI-QKFIESHKAQVRFLACVSACESVELSGVID 113  
QY 111 HLFAMAAEAVARESQVPSALLWVEPATVLCIYYFYFNGYADE-----IDAGSDEIQPLRPP 166  
DB 114 MFCSTMDIVANEFEVPSVYVFTSSAAMGL-WFHQSLRDNFGKYVDIKDSETVLS--IP 170  
QY 167 LEQSLPTFLP-----ETPERFLMMKEKLETLGDEEKAKVLVNTFDAL 212  
DB 171 AFQNLVPGVLPMFIFNTEDGCDASLDIGKRFR-----ETKGIITFLE 217  
QY 213 PDALTAIDRYELI---GIGLIPSAFLDGGDPSSETSYGDLFEKSENNCWEWLDTRPK 268  
DB 218 SHATESLDTETIPVTVGPIL-----GPK-----GSSIESLETEKILKWLDMQPE 264  
QY 269 SSVYVYSGVSLRFPPKAQMEIEIGKLLACGRPFLMWIRE-----QKNDGDEEE--EELSC 322  
DB 265 KSVVFLFCGSLGHFGAQAQVKEIAVALEGSGHRLWSLRKPPPLGRFEGPGEVENEEVLP 324  
QY 323 IGEKK---MGKIVSWCSQLEVLAIHPALGCFVTHCGWNSAVESLSGCVVAVPQWFDQ 379  
DB 325 EGFLEERTANTGMVIGWAPQTAVLSHSAVGGFVSHCGWNSLTLESIFWGPVATWPLFABQQ 384  
QY 380 TNAKLIEDAWGTGVVRMNN---EGGVDGSEIERC-----VEMVMDGGEKSKLVRENAI 430  
DB 385 MNATLVELGLAVEVRMNDYKKYKNDPADELVRADVIEEKIKILMD---PENGIRKKVK 441  
QY 431 KWKTTLAREAMGEDGSLKLNLAFLHQVAR 459  
DB 442 EMKEKSLAVEEGSSSSASLAKDFINDVIK 470

## RESULT 13

R49245  
 ID R49245 standard; Protein: 471 AA.  
 DT 08-AUG-1994 (first entry)  
 DE Anthocyanidin-3-glucoside rhamnosyltransferase (3RT).  
 KW Anthocyanidin-3-glucoside rhamnosyltransferase;  
 KW glucosyltransferase; inflorescence; flowering plants;  
 KW transgenic plant; Petunia hybrida.  
 OS Petunia hybrida.  
 PN WO9403591-A.  
 PD 17-FEB-1994.  
 PF 30-JUL-1993; AU0387.  
 PR 30-JUL-1992; AU-003846.  
 PA (ITFL-) INT FLOWER DEV PTY LTD.  
 PI Brugliera F, Holton TA;  
 DR WPI: 94-065680/08.  
 DR N-PSDB: Q56241.  
 PT Nucleic acid encoding glycosyltransferase enzymes - used for  
 PT producing transgenic plants with altered inflorescence properties  
 PT including modified petal colours  
 PS Claim 24; Page 52-55; 76pp; English.  
 CC The DNA encoding the anthocyanidin-3-glucoside rhamnosyltransferase  
 CC is used in the construction of a vector which can then be used  
 CC to transform plants. The transgenic plants are then capable of  
 CC expressing the anthocyanidin-3-glucoside rhamnosyltransferase which  
 CC confers altered inflorescence properties including modified petal  
 CC colours.  
 SQ Sequence 471 AA;

Query Match 13.5%; Score 326.5; DB 1; Length 471;  
 Best Local Similarity 26.3%; Pred. No. 5.7e-26;  
 Matches 124; Conservative 75; Mismatches 189; Indels 83; Gaps 17;

QY 6 VLLATPAGQHINPALQAKRLKAGTDTVTFTSVVWRRMANTASAAAGNPPGLDFVAF 65  
 DB 12 VVNFPPFAFGHLSPPVQLANKLSSYGVKVSFTASGNASRVKSLNSA----PTTHIVPL 67  
 QY 66 SDGYDGLKPCGDGKRYMSEMKARGSEALR-----NLLNNHNDVTFVYVSHLFAW 115  
 DB 68 TLPHVEGLPP--GAESTAELTPASAEELKVALDLMPQIKTLHLKPHFVLFDFPAQEW 124  
 QY 116 AAEVAR---ESQVPSALLWVEPATVLCIYFFNGYADEIDAGSDEIQLPRLPLEQRS 171  
 DB 125 LPMKANGLGKITYYSVVVALSTAFITC-----PARVLPKYPSPLEDMK 169  
 QY 172 LPTFLLPET-----PERFRLMMK-----EKLET-LDGEKAKVLVNTFDAL 212  
 DB 170 KPPLGFPQTSVTSVRTFEARDLYFKSPHNGTLYDRIQSLRG--CSAILAKTCSQME 227  
 QY 213 PDALTAID---RYELIGIGLIPSAFLDGDPSSETSYGGDLFEKSENNCVENWLDTPKPS 269  
 DB 228 GPYIKYVEAQFNKPVFLIGPVVP-----DPP-----SGKLEK-----WATWLNKFE 271  
 QY 270 SVVYVSFGSVLAPFPRAQMEIECKGLLACGRPFLMWIREQKNDGDEE-----EELSCIG 329  
 DB 272 TVIYCSFGSETFLTDQVKELALGTEQTLGPFLLVNFPPANVDVSAELNRALPEGLERV 331  
 QY 330 ---GKIVS-WCSQLEVLHAPALGCFVTHGWNASVSESCGVPVAVPQWFDQTTNAKLI 385  
 DB 332 KDKGIHSGWVQOQHLLAHSSVGVCHAGFSVTEALVNDQVWLPQKQDQILNAKLV 391  
 QY 386 EDAGTGVVRVNRNMGVGDGSE-IERCVMVMDGGEK--SKLVRENAIKWK 433  
 DB 392 SGDMEAGVEINRDEGDYFGKEDIKEAVEKVMVDEKEPKGLIRENQKKW 442

## RESULT 14

R26154  
 ID R26154 standard; Protein: 534 AA.

R26154;  
 DT 27-JAN-1993 (first entry)  
 DE HUG-Br2.  
 KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;  
 KW monoglucuronide; diglucuronide.  
 OS Homo sapiens.  
 FH Key  
 FT region 12..22  
 FT /note="putative membrane-insertion signal"  
 FT 492..508  
 FT /note="putative membrane-anchoring peptide"  
 FT modified\_site 348  
 FT /note="predicted Asn-linked glycosylation site"  
 FT misc\_difference 282..285  
 FT /note="residues encoded by TGCACACGGGAAG !"  
 PN WO9212987-A.  
 PD 06-AUG-1992.  
 PF 10-JAN-1992; U00282.  
 PR 10-JAN-1991; US-639453.  
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.  
 PI Owens JS, Ritter JK;  
 DR WPI: 92-284593/34.  
 DR N-PSDB: Q27369.  
 PT Isolated gene locus UGT1, DNA segments and diagnostic probes -  
 PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome  
 PT types I and II  
 PS Disclosure; Fig 9A-I; 99pp; English.  
 CC Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have  
 CC been isolated. They are referred to as HUGBr1 (Q27369) and HUGBr2  
 CC (Q27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,  
 CC upon expression individually in COS-1 cells, encode isoforms that  
 CC catalyse the formation of the two bilirubin monoglucuronides and  
 CC the diglucuronide.  
 CC The cDNAs contain identical 3' ends (1469 bp in length) to each  
 CC other and to that of the human phenol transferase cDNA, HUGP1  
 CC (Harding et al., Proc. Natl. Acad. Sci. USA 85:8281 (1988)).  
 CC In contrast, they have unique 5' ends.  
 SQ Sequence 534 AA;

Query Match 6.0%; Score 145.5; DB 1; Length 534;  
 Best Local Similarity 25.2%; Pred. No. 1.1e-06;  
 Matches 66; Conservative 36; Mismatches 109; Indels 51; Gaps 9;

QY 214 DALTAIDRYELIGIGLIPSAFLDGDPSSETSYGGDLFEKSENNCVENWLDTPKSSVY 273  
 DB 203 DHMTFQVRVNM-LYPALSYCHTFSAPYASLASELFQRE-----VTVDLLSSASWL 256  
 QY 274 VSFSGSVLRFPPKAQMEI---IGKGLLACGRPFLMWIREQKNDGDEE-----EELSCIG 324  
 DB 257 FRSDFKVDYPRIMPNNVFIGGNCANGKPLSQEFAYINASGEHGIIVVFSLESWYSEIP 316  
 QY 325 ELKMK-----GKI-----VSMCSQLEVLHAPALGCFVTHCGW 356  
 DB 317 EKKAMAIALGAKIPQTVLWRYTGTTPSNLANNTILVKWL PQNDLLGHPMTAFITHAGS 376  
 QY 357 NSAVESLSCGVPVAVPQWFDQTTNAKLI EDAGTGVVRVNRNMGVGDGSEIERCVENW 416  
 DB 377 HGVIYSEICNGVPMVMPFLGDDQMDNAKRMETK-GAGVTLNLYE---MTSEDLNQA KAVI 432  
 QY 417 DGEKSKLVRENAIKWKTARE 438  
 DB 433 N-----DKSKENIMRSLSLHKD 450

## RESULT 15

R26153  
 ID R26153 standard; Protein: 533 AA.  
 AC R26153;  
 DT 27-JAN-1993 (first entry)  
 DE HUG-Br1.  
 KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;  
 KW monoglucuronide; diglucuronide.





**THIS PAGE BLANK (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 14:38:58 ; Search time 59.57 seconds  
(without alignments)  
118.372 Million cell updates/sec

Title: US-09-147-955-2

Perfect score: 2421  
Sequence: 1 MVRRLVLTAPPAQHINPA.....GEGSSLKLNLAFLHGVARA 460

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	762.5	31.5	471	2	US-08-466-583-2
2	762.5	31.5	471	4	PCT-US95-07820-2
3	409.5	16.9	471	3	US-09-106-464-2
4	394.5	16.3	131	2	US-08-466-583-5
5	394.5	16.3	131	4	PCT-US95-07820-5
6	391	16.2	488	2	US-08-797-226-2
7	245.5	10.1	131	2	US-08-466-583-6
8	245.5	10.1	131	4	PCT-US95-07820-6
9	220	9.1	63	2	US-08-466-583-8
10	220	9.1	63	4	PCT-US95-07820-8
11	166.5	6.9	529	4	PCT-US92-00282-7
12	160.5	6.6	531	4	PCT-US92-00282-6
13	153.5	6.3	534	4	PCT-US92-00282-4
14	148.5	6.1	531	4	PCT-US92-00282-5
15	148.5	6.1	533	4	PCT-US92-00282-3
16	121.5	5.0	506	5	5180581-2
17	120	5.0	56	2	US-08-466-583-4
18	120	5.0	56	4	PCT-US95-07820-4
19	109	4.5	414	2	US-08-750-524-1
20	107	4.4	74	4	PCT-US92-00282-24
21	106	4.4	58	2	US-08-466-583-9
22	106	4.4	58	4	PCT-US95-07820-9
23	105.5	4.4	409	2	US-08-924-254-2
24	105.5	4.4	409	3	US-09-120-249-2
25	102	4.2	52	2	US-08-466-583-7
26	102	4.2	52	4	PCT-US95-07820-7
27	100.5	4.2	1114	2	US-08-576-626A-31
28	100	4.1	421	2	US-08-576-626A-53

29	89	3.7	399	1	US-08-096-623A-20	Sequence 20, Appl
30	87.5	3.6	781	1	US-08-280-690-2	Sequence 2, Appl
31	87	3.6	1142	3	US-08-904-871-12	Sequence 12, Appl
32	87	3.6	1390	2	US-08-770-544-2	Sequence 2, Appl
33	86.5	3.6	431	1	US-07-783-705A-2	Sequence 2, Appl
34	86	3.6	2544	2	US-08-576-626A-32	Sequence 32, Appl
35	85	3.5	547	1	US-08-473-981A-6	Sequence 6, Appl
36	85	3.5	547	2	US-08-474-087-6	Sequence 6, Appl
37	84.5	3.5	410	3	US-08-858-876A-2	Sequence 2, Appl
38	84.5	3.5	834	1	US-07-977-434-10	Sequence 10, Appl
39	84.5	3.5	834	1	US-08-073-384C-6	Sequence 6, Appl
40	84.5	3.5	834	1	US-08-254-359A-6	Sequence 6, Appl
41	84.5	3.5	834	1	US-08-384-490-31	Sequence 31, Appl
42	84.5	3.5	834	1	US-08-483-043-6	Sequence 6, Appl
43	84.5	3.5	834	1	US-08-459-383-31	Sequence 31, Appl
44	84.5	3.5	834	1	US-08-458-819-10	Sequence 10, Appl
45	84.5	3.5	834	1	US-08-481-238-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1  
US-08-466-583-2  
; Sequence 2, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szczylowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: and Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/466,583  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/265,427  
FILING DATE: 24-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 11-94A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
TELEX: 49617824  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 471 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-466-583-2

Query Match 31.5%; Score 762.5; DB 2; Length 471;  
Best Local Similarity 39.1%; Pred. No. 1.3e-71;  
Matches 187; Conservative 70; Mismatches 182; Indels 39; Gaps 16;



FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Spector, Eric S.  
REGISTRATION NUMBER: 22495  
TELEPHONE: 703-415-1500  
TELEFAX: 703-415-1508  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 471 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-106-464-2

Query Match 16.9%; Score 409.5; DB 3; Length 471;  
Best Local Similarity 28.7%; Pred. No. 1.6e-34;  
Matches 142; Conservative 65; Mismatches 181; Indels 107; Gaps 21;

QY 11 FP--AOGHINPQAFKRLKAGTDVTFET-----SVYAWRRMAN-----TAS 51  
DB 10 FPMAGCHMPTDMAKLVACRGVKATITTPLESVFSKAIERNKHUGLGEIDIRLLKFP 69  
QY 52 AAAGNPP-----GLDFVAFSDYDDGLKPGDGKRYMSEMKARGSEALRNLLNHHVDVTFV 107  
DB 70 AKENDLPEDCDERLDVPSDDKLPNFKAAAMMKDEFE-----LIGECRDPCL 117  
QY 108 VYSHLFAAAEAVERESQVALLWVEPATVICIYFYFNGYADEIDAGSDEIQLPLPPL 167  
DB 118 VSDMFLPTWTTSDAAKFSIPRIVFGTSTYFALCV-----GDSIR--RNKPF 160  
QY 168 EQRS--LPTFLPETPERFERLMKE--KLETLD-----GREKAK--VLVNTF 208  
DB 161 KNVSDTETVDFDFPHEIRLRTQTSPFQSDDETGMAPMIRKAVRESDAKSTGVFNFSF 220  
QY 209 DALEPDAITADRY-ELIG-----IGLIPSAFLDGGDPSETSYGGDLFEKS--EENNCV 260  
DB 221 YELED--YVEHYTKVVGKRNKAIGPLSLC-----NRDIEYKABGRKSSIDEHACL 270  
QY 261 EWLDTRPKSSVYVYSGSVLRFPAQMEIEIGKLLACGRPFLMWIRQKND-----DGESEE 317  
DB 271 KWLDSKSSIVYVCGSTADFTTAQOELAMGLEASGODFIWVIRTNEDWLPEGPEE-- 329  
QY 318 EELSCIGELAKMKKIV-SWCSQLEVLAPALGCFVTHCGWNSAVESLSCGVPVAVPQWF 376  
DB 330 -----RTEKGLIIRGNAPQVILIDHEAIGAFVTHCGWNSLLEGISAGVPLTWPVFA 382  
QY 377 DQTNNAKLIEDAWGTGVRV-----RMNEGQVDSIERCEVMYMDGGEKSKLVRENAIK 431  
DB 383 EQFFNEKLVTEVMSRGAGVGSKQKRTASGVKREATAKAIRVM-ASEETEGFRSRAKE 441  
QY 432 WKTLARAMEGDSS 446  
DB 442 YKEMAREABEGSS 456

RESULT 4  
US-08-466-583-5  
Sequence 5, Application US/08466583  
Patent No. 5919998  
GENERAL INFORMATION:  
APPLICANT: Bandurski, Robert S.  
APPLICANT: Szerszen, Jędrzej B.  
APPLICANT: Szczygłowski, Krzysztof  
TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
TITLE OF INVENTION: and Plant Growth.  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee & Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado

COUNTRY: United States of America  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,583  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/265,427  
FILING DATE: 24-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Feiber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 11-94A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
TELEX: 49617824  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: Internal  
US-08-466-583-5

Query Match 16.3%; Score 394.5; DB 2; Length 131;  
Best Local Similarity 53.7%; Pred. No. 7.6e-34;  
Matches 73; Conservative 19; Mismatches 37; Indels 7; Gaps 2;

QY 259 CVEWLDTRPKSSVYVYSGSVLRFPAQMEIEIGKLLACGRPFLMWIRQKNDGDEEEE 318  
DB 1 CTRWLDTRPKORSVAIVYSGSLASGNQAKRELARGLLAAGKPLWVVRN-----SDEHQV 55  
QY 319 ELSGICELKMKMG--KIVSWCSQLEVLAPALGCFVTHCGWNSAVESLSCGVPVAVPQWF 376  
DB 56 PRYLAETATGAAMVVPWCQDLVLAHPAVGCFVTHCGWNSLLEALSFGVPMVAMALMT 115  
QY 377 DQTNNAKLIEDAWGTG 392  
DB 116 DQPTNARNVELAMGAG 131

RESULT 5  
PCT-US95-07820-5  
Sequence 5, Application PC/TUS9507820  
GENERAL INFORMATION:  
APPLICANT: Bandurski, Robert S.  
APPLICANT: Szerszen, Jędrzej B.  
APPLICANT: Szczygłowski, Krzysztof  
TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
TITLE OF INVENTION: and Plant Growth.  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee & Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: United States of America  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/07820  
; FILING DATE: 19-JUN-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94B PCT  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 131 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
PCT-US95-07820-5

Query Match 16.3%; Score 394.5; DB 4; Length 131;  
Best Local Similarity 53.7%; Pred. No. 7.6e-34;  
Matches 73; Conservative 19; Mismatches 37; Indels 7; Gaps 2;  
  
QY 259 CVELDTPKSSVYVSGSVLPFPKQMEIEGKLLACGRPFLMWIREQKNDGEEEEE 318  
DB 1 CTKWLDTPDRSVAYVSGSLASLNAQKEELARGLAAGKFFLWVRA-----SDEHOV 55  
  
QY 319 ELSCIGELKKMG--KIVSMCSQLEVLHPALGCFVTHCGWNSAVESLSGVPVAVPQWF 376  
DB 56 PRYLLAEATATGAAMVWPQDVLHPAVGCFVTHCGWNSLLEALSFGVPVMAWLWT 115  
  
QY 377 DQTTNAKLIEDAWGTG 392  
DB 116 DQPTNARNVELAWGAG 131

RESULT 6  
US-08-797-226-2  
; Sequence 2, Application US/08797226  
; Patent No. 5959180  
; GENERAL INFORMATION:  
; APPLICANT: MOEHS, CHARLES P  
; APPLICANT: ALLEN, PAUL V  
; APPLICANT: ROCKHOLD, DAVID R  
; APPLICANT: STAPLETON, ANDREW  
; APPLICANT: GARBARINO, JOAN E  
; APPLICANT: FRIEDMAN, MENDEL  
; APPLICANT: BELKNAP, WILLIAM R  
; TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE  
; TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE  
; TITLE OF INVENTION: GLYCOALKALOIDS IN SOLANACEOUS PLANTS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NANCY J. PARSONS  
; STREET: 800 BUCHANAN ST.  
; CITY: ALBANY  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94710  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/797,226  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PARSONS, NANCY J  
; REGISTRATION NUMBER: 40,364  
; REFERENCE/DOCKET NUMBER: 0011.97  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 559-5731  
; TELEFAX: (510) 559-5777  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 488 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-797-226-2

Query Match 16.2%; Score 391; DB 2; Length 488;  
Best Local Similarity 23.9%; Pred. No. 1.5e-32;  
Matches 125; Conservative 89; Mismatches 186; Indels 122; Gaps 20;  
  
QY 6 VLLATPPAQGHINPALQFAKRLKAGTDVTFSTVYAMRRMANT----- 49  
DB 13 VLFLPFLSAGHFPLVNAARLFASRGVKATILUTPHNALLPRSTIDDDVRISGFPISIVT 72  
  
QY 50 ---ASAAAGNPGDLDFVAFSDGYDGLAPCGDGKRY--MSENKARGSEALRLLNNHNDV 104  
DB 73 IKFPAEVLGPEGI-----ESFNSATSPMPHKIFYALSLLQKPMEDKIRELRPD---- 122  
  
QY 105 TPVWYSHL-FANAAEVARESQVPSALLWVEPATVLCIYVFY-FNGVADEIDAGSDRIQLP 162  
DB 123 --CIFSDMWFPTVDIADDELHPRIL--YNLSAYWCYSIMHNLKVYRPHKQNLDESQ--- 176  
  
QY 163 RLPPLEQRSLPTFLPETPERFLMMKEKLETL-----DGEKAKYL 204  
DB 177 -----SFVVPGLPDEIKFKLSQLTDDLKSDQKTVFDELLEQVEDSEERSYGI 225  
  
QY 205 V-NTFDALPDALTADRYELI-----GIGPLISAFLDGDPSETSYGGDLFEK---S 254  
DB 226 VHDTFEYELP---AYVDYQKLLKPKCMHFGPL-----SHFASKIRSKELIS 269  
  
QY 255 EENN---CVELDTPKSSVYVSGSVLPFPKQMEIEGKLLACGRPFLMWIREQKND 311  
DB 270 EHNNEIVIDLWLNQPKSVLYVSGSMARFPESQLNEIAQALDASNVDFIVLRP---- 325  
  
QY 312 DGEEEELSCIGELKKMGK---IVSMCSQLEVLHPALGCFVTHCGWNSAVESLSGCV 367  
DB 326 --NEETASWLPVGNLEDKTKGLYIKGWVPQLTMEHSATGCGTNSVLEAITFGV 383  
  
QY 368 PVVAVPQWFDQTTNAKLIE-----DANGTGVVRVMEGGVGDGSEIERCIVE--MY 415  
DB 384 PMITWPTADQFYNKVEVRLGKIGKIDVWNEGIEI---TGPVIESAKIREATERLMI 440  
  
QY 416 MDGGEKSLVRENATKWKTLAREMGEDGSSLLKNLNAFLRQV 457  
DB 441 SNGSEIINIRVRVAMSKMAQNATNEGSSWNNLTALIQHI 482

RESULT 7  
US-08-466-583-6  
; Sequence 6, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szeszen, Jędrzej B.  
; APPLICANT: Szczygłowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: and Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.

APPLICANT: Szerszen, Jędrzej B.  
APPLICANT: Szczygłowski, Krzysztof  
TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
TITLE OF INVENTION: Genetic Control of Plant Growth.  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee & Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: United States of America  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,583  
FILING DATE: 06-JUN-1995





RESULT 13

: GENERAL INFORMATION: IDA S.  
 : APPLICANT: OWENS, JOSEPH K.  
 : TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION  
 : TITLE OF INVENTION: THEREIN.  
 :

CORRESPONDENCE ADDRESS:

STREET: 1615 L STREET, N.W.

STATE: D.C.

CONFIDENTIAL  
ZTP: 20036-5601

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC  
OPERATING SYSTEM: PC-DOS/MS

: CURRENT APPLICATION DATA:

AFFIDAVIT NUMBER: FC1/0392/00203  
 FILING DATE: 19920110

ATTORNEY/AGENT INFORMATION

REGISTRATION NUMBER: 265

: TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-691-2000  
TELEFAX: 202-822-0944

: INFORMATION FOR SEQ ID NO:

LENGTH: 534 amino acids

TYPE: AMINO ACID  
STRANDEDNESS: single

TOPOLOGY: protein

FCI-0392-00202-4

### Query Match

Best Local Similarity 24.2%, Fied. NO: 1.3E-07,  
Matches 51: Conservative 41: Mismatches 90: Indels 29

0v 229 PLTPSAFLDGGDPSETSYGGDLFEKSEFNNOVFVWLTMTKPKSSVWVVSFCS-VI.RFPKAO 287

268 PTMPNMFICG--TNCANGKPIISOEEF-----AYTNASCFHGTVVFSICSMVSEIPEKKA 320

0v 288 FFTGKGLIACGRPEIWMTPFOKNDDGEEEEETCTGELKMKGTVSWCSOLFVLAHPAT. 347

query match  
Best Local Similarity 23.5%  
Best Local Similarity 23.5%  
Score 100.0, DB 4, Length 331,  
Pred. No. 2.8e-08:

STATION 40, WATSONVILLE  
STATION 40, WATSONVILLE

[illegible]

205 4 AMFNNHAT100  
INC VTRAPDSQEFER1V NBSQEHG1V VFSZGS 307

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

000 MVSEIFERNMELTAEADGNIFQILDMNIGIRFSN-----DARNJTLVAKWLF 334

Y 358 QDEVLAHFALGCFVIRCGWNSAVESSLSCGVFVVAVFQWFDTIINAKLIEDAWGIGVRVRM 3597

PCT-US92-00282-5

; Sequence 5, Application PC/TUS9200282

; GENERAL INFORMATION:

; APPLICANT: OWENS, IDA S.

; APPLICANT: RITTER, JOSEPH K.

; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION

; TITLE OF INVENTION: THEREIN.

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CUSHMAN DARBY &amp; CUSHMAN

; STREET: 1615 L STREET, N.W.

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20036-5601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/00282

; FILING DATE: 19920110

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: SCOTT, WATSON T.

; REGISTRATION NUMBER: 26581

; REFERENCE/DOCKET NUMBER: 91532-PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-861-3000

; TELEFAX: 202-822-0944

; TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 531 amino acids

; TYPE: AMINO ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US92-00282-5

Query Match

Best Local Similarity 6.1%; Score 148.5; DB 4; Length 531;

Matches 51; Conservative 40; Mismatches 91; Indels 29; Gaps 7;

QY 229 PLIPSAFLDGGDPSETSYGGDLFEKSEENNCVWLDTKPKSSVYVYVSGS-VLRFPKAOH 287

Db 265 PMPNNMVFVGG--INCKRKDLUSQFE-----AYINASGEHGIIVFSLGSMVSEIPEKKA 317

QY 288 EETGKGLACGRPFLLMIREQKNDGEEELSCIGELKMGKIVSWCSQLEVLARPAL 347

Db 318 MAIADALGNPQTVLWRYTGTSPN-----LANNTILVKWLPQNDLLGHPMT 364

QY 348 GCFVTHCGNWSAVESLSCGVPVAVPQWFDQTTNAKLIEDAGTGVVRMNEGGVDSGE 407

Db 365 RAFITHAGSHGVYESICNGVPVMPVMPFLFGDQMDNNAKRMETK-GAGVTLNVLE---MTSED 420

QY 408 IERCVMVMDGGEKSKLVRENAIKWKTARE 438

Db 421 LENALKAVIN---DKSYKENIMRLSSLHKD 447

RESULT 15

PCT-US92-00282-3

; Sequence 3, Application PC/TUS9200282

; GENERAL INFORMATION:

; APPLICANT: OWENS, IDA S.

; APPLICANT: RITTER, JOSEPH K.

; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION

; TITLE OF INVENTION: THEREIN.

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CUSHMAN DARBY &amp; CUSHMAN

; STREET: 1615 L STREET, N.W.

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20036-5601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/00282

; FILING DATE: 19920110

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: SCOTT, WATSON T.

; REGISTRATION NUMBER: 26581

; REFERENCE/DOCKET NUMBER: 91532-PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-861-3000

; TELEFAX: 202-822-0944

; TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 533 amino acids

; TYPE: AMINO ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US92-00282-3

Query Match

Best Local Similarity 6.1%; Score 148.5; DB 4; Length 533;

Matches 50; Conservative 39; Mismatches 83; Indels 49; Gaps 7;

QY 229 PLIPSAFLDGGDPSETSYGGDLFEKSEENNCVWLDTKPKSSVYVYVSGS 278

Db 267 PMPNNMVFVGG-----INCLHQNPDSQFEAYINASGEHGIIVFSLGS 309

QY 279 -VLRFPKAOHEEIGKGLACGRPFLLMIREQKNDGEEELSCIGELKMGKIVSWCS 337

Db 310 MYSEIPEKKAMAIADALGNPQTVLWRYTGTSPN-----LANNTILVKWLP 356

QY 338 QLEVLARPALGCFVTHCGNWSAVESLSCGVPVAVPQWFDQTTNAKLIEDAGTGVVRM 397

Db 357 QNDLLGHPMTAFITHAGSHGVYESICNGVPVMPVMPFLFGDQMDNNAKRMETK-GAGVTLNV 415

QY 398 NEGGVDSGEIERCVMVMDGGEKSKLVRENAIKWKTARE 438

Db 416 LE---MTSEDLENALKAVIN---DKSYKENIMRLSSLHKD 449

Search completed: August 1, 2000, 19:51:30

Job time: 18752 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2000, 21:35:38 ; Search time 9137.45 Seconds  
(without alignments)  
294.342 Million cell updates/sec

Title: US-09-147-955-1  
Perfect score: 1507  
Sequence: 1 gaaaatttcacacaaaatgg.....ttataatttagtaacaaaa 1507

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 972840 seqs, 892348106 residues  
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pr1: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_sts: \*  
14: gb\_sy: \*  
15: gb\_un: \*  
16: em\_fun: \*  
17: em\_hum1: \*  
18: em\_hum2: \*  
19: em\_in: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_sy: \*  
29: em\_un: \*  
30: em\_vi: \*  
31: gb\_htg1: \*  
32: gb\_htg2: \*  
33: gb\_in1: \*  
34: gb\_in2: \*  
35: em\_ba1: \*  
36: em\_ba2: \*  
37: em\_hum3: \*  
38: em\_hum4: \*  
39: gb\_pr4: \*  
40: gb\_htg3: \*  
41: gb\_htg4: \*  
42: gb\_htg5: \*  
43: gb\_htg6: \*

44: gb\_htg7: \*  
45: em\_hta1: \*  
46: em\_htg2: \*  
47: em\_htg3: \*  
48: em\_hum5: \*  
49: gb\_pl3: \*  
50: gb\_pr5: \*  
51: gb\_htg8: \*  
52: gb\_htg9: \*  
53: gb\_htg10: \*  
54: gb\_htg11: \*  
55: gb\_htg12: \*  
56: gb\_htg13: \*  
57: gb\_htg14: \*  
58: gb\_in3: \*  
59: gb\_htg15: \*  
60: gb\_htg16: \*  
61: gb\_htg17: \*  
62: em\_htg4: \*  
63: em\_htg5: \*  
64: em\_htg6: \*  
65: em\_htg7: \*  
66: em\_hum6: \*  
67: gb\_htg18: \*  
68: gb\_htg19: \*  
69: gb\_htg20: \*  
70: gb\_htg21: \*  
71: gb\_htg22: \*  
72: gb\_htg23: \*  
73: gb\_htg24: \*  
74: gb\_htg25: \*  
75: gb\_htg26: \*  
76: gb\_htg27: \*  
77: gb\_htg28: \*  
78: gb\_htg29: \*  
79: gb\_htg30: \*  
80: gb\_htg31: \*  
81: gb\_v11: \*  
82: gb\_v12: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	Query Match %	Length	DB ID	Description
1	1507	100.0	1507	7	AB013596	AB013596 Perilla f
2	1329.4	88.2	1458	7	AB013597	AB013597 Perilla f
3	700.2	46.5	2043	7	AB013598	AB013598 Verbena h
4	455.4	30.2	1864	7	AB027455	AB027455 Petunia x
c 5	300	19.9	198944	8	ATCHRIV38	AL161538 Arabidops
c 6	300	19.9	200576	7	ATFCA0	297335 Arabidops
7	283	18.8	1594	7	AB000623	AB000623 Nicotiana
c 8	222.8	14.8	84203	49	AC005106	AC005106 Genomic s
c 9	222.8	14.8	103223	8	AC007153	AC007153 Arabidops
c 10	184.4	12.2	195068	8	ATCHRIV42	AL161542 Arabidops
c 11	184.4	12.2	197419	8	ATCHRIV41	AL161541 Arabidops
c 12	184.4	12.2	205065	7	ATFCA4	297339 Arabidops
13	181.2	12.0	1589	8	ATU81293	U81293 Arabidops
14	160.8	10.7	99188	49	AC006533	AC006533 Arabidops
15	159.6	10.6	1513	5	A62523	A62523 Sequence 28
16	159.6	10.6	1513	5	A72396	A72396 Sequence 28
17	149.2	9.9	1588	5	A62526	A62526 Sequence 31
18	148.4	9.8	1459	5	A62525	A62525 Sequence 30
19	145.2	9.6	1969	5	A62529	A62529 Sequence 34
c 20	137.2	9.1	123358	8	AC002396	AC002396 Arabidops
21	132	8.8	80818	7	AB019232	AB019232 Arabidops
22	131.8	8.7	1416	7	MZEIAGLO	L34847 Zea mays IA
23	131.8	8.7	94786	49	AC002333	AC002333 Arabidops
24	127.4	8.5	1440	49	AF190634	AF190634 Nicotiana



```

Db  841 GTTTGGGAGCGTTTGTAGGTTTCCAAAGCGCACAAATGAAGAGATTGGAAAGGCGCTATT 900
QY  901 agcctcggaagcgctgttttatgtatgatacagagacacaaataacacacgcgaaga 960
Db  901 AGCTCGGGAAGCGCTTTTATGTGATGATACGAGAACAGAAGTAACACGCGCGAAGA 960
QY  961 agagaagaagagttgagttgcatggtgggaattgaaaaaattggggaataatagtttcgtg 1020
Db  961 AGAAGAAGAAGACATTGATTCGATTGGGAATTGAAAAAATGGGGAATAGTTTCGTG 1020
QY  1021 gtgctcgagttgagagttcttgcgcacctgogttgggagttgttctgaogcattgttg 1080
Db  1021 GTGCTCGAGTTGAGGTTCTTGGCGCACCCCTGCGTTGGGATGTTTCGTGACGCATTGTGG 1080
QY  1081 gtggaactcgctgtgagagcttgatcggtgttcccggtgtgtgctgcgcgaagt 1140
Db  1081 GTGGAATCGCCTGTGAGAGCTTGTAGTTGCGGGTTCGCGTGTGCGGTGCGCGCAGTG 1140
QY  1141 gtttgatcagacgacgaatgcgaagctgattgaggaatgcgtgggggacaggggtgagagt 1200
Db  1141 GTTTGATCAGACGACGAATCGGAAGCTGATTGAGGATGCGTGGGGGACAGGGGTGAGAGT 1200
QY  1201 gagaatgaatgaaggggtggttggatgatctgagatgagagaggtgtgtggagatggt 1260
Db  1201 GAGAAATGAATGAAGGGGTGGGTGTGATGGATCTGAGATCAGAGAGGTGTGTGGAGATGTT 1260
QY  1261 gatgatgggggtgagagagcaactagtagagagaaaaatgccataaaatggaagaattt 1320
Db  1261 GATGGATGGGGGTGAGAGAGCAAACTAGTGAGAGAAAATGCCATAAAATGGAAGACTTT 1320
QY  1321 ggcagagaagccatgggagagatgatttcactcaagaatctcaacgcttcttca 1380
Db  1321 GGCCAGAGAAGCCATGGAGAGGATGATCTTCACTCANGAATCTCAACGCCTTCTTCA 1380
QY  1381 tcaagttgacgtgcttaatacacaaaatggcttccactttcaatttcaactcaaacccg 1440
Db  1381 TCAAGTTGCGAGTCTTAATACACAAAATGGCTTTCACACTTTAATCTACTCAACACCG 1440
QY  1441 gtcaataaataatcccttccacttcttctatttccactatcacatttaattttagt 1500
Db  1441 GTTCAATAAATAATCCCTTCCACTTCTTCTATTCTACTATCATCATTTAATAATTTAGT 1500
QY  1501 aacaaa 1507
Db  1501 AACAAA 1507

RESULT 2
AB013597
LOCUS      AB013597      1458 bp      mRNA      PLN      20-MAR-1999
DEFINITION Perilla frutescens PF3R6 mRNA for UDP-glucose:anthocyanin
5'-O-glucosyltransferase homologue, complete cds.
ACCESSION AB013597
VERSION    AB013597.1 GI:4115560
KEYWORDS   Perilla frutescens:anthocyanin 5-O-glucosyltransferase homologue.
SOURCE     Perilla frutescens (variety:crispa, cultivar:Shikun) leaf cDNA to
           mRNA
ORGANISM   Perilla frutescens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
            Asteridae; Gentiananae; Lamiales; Lamiaceae; Perilla.
REFERENCE  1 (sites)
AUTHORS    Yamazaki,M., Gong,Z., Fukuchi-Mizutani,M., Fukui,Y., Tanaka,Y.,
            Kusumi,T. and Saito,K.
TITLE      Molecular cloning and biochemical characterization of a novel
            anthocyanin 5-O-glucosyltransferase by mRNA differential display
            for plant forms regarding anthocyanin
JOURNAL    J. Biol. Chem. 274 (11), 7405-7411 (1999)
MEDLINE    99167509
REFERENCE  2 (bases 1 to 1458)
AUTHORS    Yamazaki,M., Saito,K. and Gong,Z.
TITLE      Direct Submission
JOURNAL    Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami

```

```

FEATURES             source
    source            1..1458
                        /organism="Perilla frutescens"
                        /cultivar="Shikun"
                        /variety="crispa"
                        /db_xref="taxon:48386"
    gene              29..1360
                        /tissue_type="leaf"
    CDS                29..1360
                        /gene="PF3R6"
                        /gene="PF3R6"
                        /codon_start=1
                        /product="UDP-glucose:anthocyanin 5-O-glucosyltransferase
                        homologue"
                        /protein_id="BAA36422.1"
                        /db_xref="GI:4115561"
                        /translation="MWRRRVLLATFFPAQGHINPALQFAKRLLLKAGTDVTFYVYVNR
                        RMATASAAAGNPGLDFVAFSDGYDDLKPPGGDKRIMSEMKGARGEALRNLILLND
                        DVTFVYSHLFAMAAEVARLSHVPTALLWEPATVLCIYHFYFNGYADEIDAGSNEIQ
                        LPRPLSEQRSLPTFLPATPERFLMKKLETLDEKAKVLTNTFDALEPDALTA
                        IDRVELIGLIPSAFLDGEDPSETSYGDDLPEKSENNCVLWNSPKSVVYVF
                        GSVLRPPKAOMERIGKLLACGRPFLLMI REQNDKDEEEBELSCIGELKKMGKI
                        VSWCSQLEVLAHFALPGCFVTHCGWNSAVESLSCGIPVAVPOHFDTNKLIEDANG
                        TGVVRNMEGGVGGCBEIERCEVEMVMDGGDKTLVRENAIRKWLARQAMG"
    polyA_site        1458
                        /note="16 a nucleotides"
    BASE COUNT        353 a 348 c 438 g 319 t
    ORIGIN
Query Match      88.2%; Score 1329.4; DB 7; Length 1458;
Best Local Similarity 95.4%; Pred. No. 5.4e-280;
Matches 1403; Conservative 0; Mismatches 36; Indels 32; Gaps 2;
QY  2 aaatttccacaaaatggtccgcgcgctgtgctgtagcaagcttcttcgcgaagc 61
Db  14 AAAATTTCCACAAAATGGTCCGCGCGCGTGTGCTAGCAACGTTTCCGCGCAAGGC 73
QY  62 cacataaatccgcctcccaattcgccaagagactcctaaaagccggcactgaactaca 121
Db  74 CACATAAATCCGCCCTCCAAATTCGCCAAGAGACTCTCTAAAAGCCGGCAGCTCAGC 133
QY  122 ttttcaagagcgtttatgcatggcgcgcgcatggccaacacagcctccgcgtccgga 181
Db  134 TTTTTCAGGAGCGTTTATGATGTCATGGCGCGCATGGCCAAACACAGAGCTCGCGCTGCCGA 193
QY  182 aaccacacgggcctcgacttgcgtgttccgacggtacgacgacgagggctgaagccc 241
Db  194 AACCCACCGGCGCTCGACTTCGTGCGCTTCCCGAGGCTACGACGAGGGGTGAAGCCC 253
QY  242 tgcgcgcagcgaagcgtcatatgccagatgaaagccgcgctccgagccttaaga 301
Db  254 GGCGGCGAGGGAAGCGGTACATGTCCGAGATGAAGAAGCCCGCGGCTCCGAGCCTTAGA 313
QY  302 aactcctcttcaacaacacacgacgactcgtctcgtctactccacctctttgcatgg 361
Db  314 AACCTCTTCTCAACACAGGACGAGCTACTTTGTCGCTACTTCCCACTCTCTTTCATGG 373
QY  362 gcgcgcgaggtggcgcgtgagtcaccaggtcccgagcgccctctctgtgggtcgagcccgc 421
Db  374 CGCGGGAGGTGGGCGGCTTGTTCACACGTCCGACCGCCCTCTCTGGGTGAGGCCGCC 433
QY  422 accgtgctgcatatattacttactcaacgctacgacgacgacgacgacgacgacg 481
Db  434 ACCGTGCTGTGCATATACCACTTCTACTTCAACGGCTACGACAGGAGTACGACCCGGT 493
QY  482 tccgacgaattcagctccctcggtctccacccctggagcagcgagtccttcacacctt 541
Db  494 TCCAATGAATTCAGCTCCCTCGGCTTCCATCCCTGGAGAGCGCAGTCTTCCGAGCTTT 553

```

Yamazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. & Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan  
(E-mail:mamiy@p.chiba-u.ac.jp, Tel.:+81-43-290-2905, Fax:+81-43-290-2905)



```
QY 301 -----aaacctctcttcaacaacacgacgctcagcttcgctctac 343
Db 302 GATACCTTCAGCCATATATCGCATCAAAAAGCAGCAAAATCACGTTCTGGGTGATC 361
QY 344 tccacactcttgcagggcgccggaggtgagcgctgagtcaccaggtcccgagcgccctt 403
Db 362 TCCACACCTCTTTCATGGCGGCGCAAGGTGGCGGTGAGTTCATCTCCGAGCGCGCTA 421
QY 404 ctctgggtcagcccgccaccgctgctgcataatattacttctcaacggctaagca 463
Db 422 CTCTGGATTGAGCAGCTACGGGTGAGATATATTTTATTTTATTTCAACGGCTATAGC 481
QY 464 gacgagatcagccggttcgagcaaatcca---gctccctcggttccacccttgag 520
Db 482 GACGAATCATCGGGTTCGGATGCTATTCACTTCCCGGAGGACTCCAGTCTGCC 541
QY 521 cagcgagcttcctcagaccttctcgtccggagacacccggagagattccggttgatg 580
Db 542 CAGCGTGATTTACCGTCTTCCCTTCCCTTCCACGCATGAGAGATTCCTGCTACTGATG 601
QY 581 aaggagaagctgaacctttagacggtgaagagaagcgaaagtgttgtaaacacattt 640
Db 602 AAGGAGAAATGGAACTTTAGAAAGGTGAAGAAACCTAAGGTCCTTGGTGAACAGCTTT 661
QY 641 gatgcttgagcccgatgcactcacgctattgataggtgatgagttgacggtcg 700
Db 662 GATCGTTGGAGCTGATCGCTCAAGCCCATTCATTAAGTACGAGATGATTCNAATCGG 721
QY 701 ccgttgattccctccgcttcttgacgcgcgagatccctcccaaacgtcttaccggc 760
Db 722 CCGTGTGATCTTCCGATCTTTCGACGGTAAAGATCTTCGGACAGGTCTTCGGCGGA 781
QY 761 gatcttttcgaaaaa---tcggaggagaataaactcgctgaggtgttgacacagacgg 817
Db 782 GATTGTGTCGAGAAAGGGTTCGATGACGAGATTGCTCGAATGGTTGAGCACGAATCCT 841
QY 818 aaatcttcggtggtgatgctgctggtggagcggtttgaggtttccaaaggcacaatg 877
Db 842 CGATCTTCGTTGTTACGTTTCCTCGGAAGCTTCGTTAATACGACAGATGCCAATG 901
QY 878 gaagagattgggaagggctattagctcgcggaagggcgtttttatgatgatacagaa 937
Db 902 GAAGAGATAGCAAGAGGCTGTAGATTGTGGAGGCCGCTTTTGTGGTGGTGAAGATA 961
QY 938 cagaagaatgacgacggcggaagaagaagaagatgagttgattgggaattgaaa 997
Db 962 AA-----CGAAGGAGAGAGAGGTATTGATAAGTTGTCATGGAGGAGTTGAAA 1006
QY 998 aaatggggaaaaatgcttcggtgctcgcagttgaggttctggcgccacctgcgttg 1057
Db 1007 CGAGTGGGAAAAATTGATCTTGGTGTCTCAATTGGAAGTCTCGACGATCCCTCGTTG 1066
QY 1058 ggaatgcttgacgacatttggtgggaactcggctgagagagcttgagttgcgggggtt 1117
Db 1067 GGATGTTTCGTGACACACTCGGGTGAATTCGACTCTAGAGAGTATATCTTTTCGGGTT 1136
QY 1118 ccggtggtgaggggtgcgacgagtttgatcagacacgaatacgaagctgattgaagat 1177
Db 1127 CCGATGGTGGCTTTTCGCGAGTGGTTTCGATCAAGGGACGAATGCGAAGCTGATGGAGAT 1186
QY 1178 gctgggggacaggggtgagatgagaaatgaatgaaggggtt---gggggttgatggatc 1234
Db 1187 GTGTGGAGGACGGGTGTGAGAGTGAAGACTAATGAGGAGGGTAGCGTCTGTGATGGTAT 1246
QY 1235 gagaatagaggggtgtgtgagatggtgatggtgggggttgagaaagacaaactagtga 1294
Db 1247 GAAATTAGGAGATGATTGAGGAGGTTTATGGATGGGGGAGAAAAGAGTAGGAAACTTGA 1306
QY 1295 gaaaatgccataaaatgagacatttggccagagagaacccatggagagggatgagatcttca 1354
Db 1307 GAGAGTCTGGCAAGTGAAGGATTTGGCAAGAAAAGCTATGGAGGAAGATGGATCTTCA 1366
```

```
QY 1355 ctcaagaatctcaacgctcttcttctcaacgctgctgcttaatacacaaaaatggctt 1414
Db 1367 GTTAACAACCTCAAGGCTCTTCTTGATGAGGTTGATAGTATCTAAAGACGCTAAATGAGGT 1426
QY 1415 tcc 1417
Db 1427 CCC 1429

RESULT 4
AB027455 1864 bp mRNA PLN 07-JAN-2000
LOCUS Petunia x hybrida PH1 mRNA for anthocyanin 5-O-glucosyltransferase,
DEFINITION complete cds.
ACCESSION AB027455
VERSION AB027455.1 GI:6683051
KEYWORDS anthocyanin 5-O-glucosyltransferase.
SOURCE Petunia x hybrida (cultivar:Surfinia) corolla cDNA to mRNA.
ORGANISM Petunia x hybrida
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
REFERENCE 1 (bases 1 to 1864)
AUTHORS Yamazaki,M. and Saito,K.
TITLE Anthocyanin 5-O-glucosyltransferase
JOURNAL Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 1864)
AUTHORS Yamazaki,M. and Saito,K.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) to the DDBJ/EMBL/GenBank databases. Mami
Yamazaki, Chiba University, Faculty of Pharmaceutical Sciences,
Inage-ku Yayoi-cho 1-33, Chiba, Chiba 263-8522, Japan
(E-mail:mamiy@p.chiba-u.ac.jp, Tel:81-43-290-2905,
Fax:81-43-290-2905)
FEATURES
source
1..1864
/organism="Petunia x hybrida"
/cultivar="Surfinia"
/db_xref="taxon:4102"
/tissue_type="corolla"
100..1506
/gene="PH1"
100..1506
/gene="PH1"
/codon_start=1
/product="anthocyanin 5-O-glucosyltransferase"
/protein_id="BAAB9009.1"
/db_xref="GI:6683052"
/translation="MVQPHVILTFTPAQGHINPALQPAKMLVKMGIEVTFSTSIYQAS
RMDKSTLNAPKGLNFPFSDGDFDHSKDPVFMYSOLRKCSEVTKVKKIILTCSEN
GQPTCLLYSIFILPMAAEVAREVHIIPSAALWSOPATILDIYFNHGVKAMANESEN
PNWSIQLPGLLETRDLPSFLLPYGAKGSLRVALPFPKELIDTLDATETPKILYNTF
DEVPALNAIEGTFTGIGLIPSAFLGNDPDLDAISFGDLFONSDNDEWLNSKPN
SEUYYSFSGLMNPSISQMEELIDIGRFLWIKENEKKEENKKGICKEELE
KIKGIWPCSQLEVLKHPSCFVSHGWSALESLACGVPVAVFPQWTDQMTNAKV
EDWKGVRVRIINEDGVVESEIKRCITELVMDGEGKEELRKNKAKKELAREAVKEG
GSSHKNLKAFIDDVAKGF"
BASE COUNT 603 a 318 c 369 g 574 t
ORIGIN

Query Match 30.2%; Score 455.4; DB 7; Length 1864;
Best Local Similarity 61.6%; Pred. No. 2.6e-89;
Matches 871; Conservative 0; Mismatches 491; Indels 51; Gaps 7;

QY 13 aaaaatggtccgcgcgcgctgctgctagcaacgcttctcgcgaagccacataatcc 72
Db 96 AAATATGGTCAGCCTCATGTCTTAAACAACATTTCCAGCACAAAGCCCATATTAATCC 155
QY 73 cgccctcaaatcgccaaagagactcctaaagccggaactgacgtcactttttcagag 132
Db 156 AGCACTTCAATTTGCCAAGATCTTGTCAAGATGGCATAGAGATGACATTTTCTACAAG 215
```

Qy	133	cgtttatgatgagcgccgcatgagccaaacacagcctcccgctcgaggaacccacccg	192
Db	216	CAATTTATGCCAAAGCGGTATGGATGAATAATCCATCTTAAATGC-----ACCAAAGG	269
Qy	193	ccLcgactcgtggcggtctccgacggtatgagacgagggctgaagccctgcgcgcacgg	252
Db	270	ATTGAATTTCAATTCCTATTTCCGATGGCTTTGTATGAAGGTTTTCATCATTCAAAAGACCC	339
Qy	253	gaagcgcacatgctccgagatgaagcccgcggtccgagcgcccttaagaacacctctctct	312
Db	330	TGTATTTACATGTCACACTTCGTAAATGTGGAAGTGAAACTGTCAAAAAATAATTTCT	389
Qy	313	c-----aaacaaccaagctcacgttctctactctccacacctcttgcgtgcatg	360
Db	390	CACCTGCTCTGAAATGGACAGCCTATAACTTGGCTACTTTACTTCCATTTTTCCTTCCTTG	449
Qy	361	ggcgcgaggggtggcggtgagtcacaggtcccgagcgccctctctcttgggtcgagccgc	420
Db	450	GGCAGCAGAGGTAGCAGCTGAAGTTTCACATCCCTTCTCTCTCTTCTTGGAGTCAACGAGC	509
Qy	421	cacgtgcgtgcacatatactctactctcaagcgtctacgcacgacgacgacgcgcgg	480
Db	510	AACAATATTGGACATATATTTACTTCAACTTTCATGGATATGAAGAAGCTATGGCTAATGA	569
Qy	481	ttccgcagaa-----attcagctccctcggtctccacccctggagcagcgacg	528
Db	570	ATCCAATGATCCAAATTTGTCTCACTTCAACTTCCCGGGCTTCCACTACTGGAAATCGAGA	629
Qy	529	tcttcgacacctctcgtctgcc-----ggagacacggagagatctccggttgatgatg---	580
Db	630	TCTTCCCTTCATTTTACTTTTACTTCTTATGGTGCRAAAGGGAGTCTTCGAGTTCGCAC	689
Qy	581	----aaggagaaagctgaaactttagcgggtgaagagagcggaagctgttggtgaacac	636
Db	690	ATTCAAGAATTTGATAGACACATTTAGTCTGAAACCACTCTTAGAATCTTTGTGAATAC	749
Qy	637	gtttgatcggtggagccgcatgcactcacgctattgataggatgatggttgatcgggat	696
Db	750	ATTTGATGAATTAGAGCCTGAGGCACCTCAATGCAATTCGAAGTTTAAAGTTTATGGAAT	809
Qy	697	cgggcggtgatccctccgcctctttggacgcgagagatccctccgaaacgctctacgg	756
Db	810	TGACCGCTTGATTTCTCTCTCTCTTCTGGTGGAAATGACCCCTTTAGATGCTTCATTTGG	869
Qy	757	cggcgcactcttcgaaaaactcgagagaaataactgcgtggagtggttgagcacgacgac	816
Db	870	TGTTGATCTTTTTCAAAATCAAAATGACTATA-----TGAATGGTTAAACTCAAGACC	923
Qy	817	gaaatctcgggtggtgatgtcgctttgggagcgtttgagggtttccaaaggcacacaa	876
Db	924	AAATTCATCAGTTGTTTATATATCTTTTGGGAGTCTAATGAATCCATCTATTAGCCAAAT	983
Qy	877	ggaagagattggaaaggcctattagctgcggaaggcgctttttatgga tga tacaaga	936
Db	984	GGAGGAGATATCAAAAGGGTTGATAGACATAGACAGGCGCTTTTATGGGTGATTAAGA	1043
Qy	937	acagagaatgacgcgcggaagagaagaagagtgatgagtcattggggaattgaa	996
Db	1044	AAATGAAAAGGCCA----AGAAAGAGAGAAATAAANGCTTGTGTGATTTGAAGAAATTTGGA	1100
Qy	997	aaaaatggggaaaaatggttcgtggtcgcagttgagggtctctggcgaccctgcgtt	1056
Db	1101	AAAAATAGGAAAAATAGTTCCATGGTTTCACAACTTGAAGTTCTTAAACATCATCTTT	1160
Qy	1057	ggatgttctcgtgacgcattgtgggtggaactcgcgctgtgagagccttgagttgcggggt	1116
Db	1161	AGGATGTTTTGTTTCTCATTTGTGGATGGAAATTCAGCCTTAGAGAGTTTAGCTTGTGAGT	1220
Qy	1117	tcgggtggtggcggttcccgagtggttgatcacagacgacgaatgcgaagctgattgaga	1176
Db	1221	GCCAGTTGTGGCATTTTCTCAATFGACAGACATCAATAGCAATGCCAAACAAAGTTGAAG	1280
Qy	1177	tdcgtgggggacaggggttdagagtdaagaatgaatgaacgggggttgggttgaatgactga	1236

[illegible]



exon	/number=2 complement(9562.. .9654)	intron	/number=12 complement(12306.. .12438)
gene	/number=1 9562.. .15425 /gene="AT4g14070"	exon	/number=12 complement(12439.. .12554)
CDS	complement(join(9562.. .9654,9750.. .9807,9883.. .9998, 10097.. .10176,10276.. .10435,10594.. .10731,11065.. .11145, 11247.. .11300,11383.. .11541,11653.. .11758,11882.. .11949, 12156.. .12305,12439.. .12554,12626.. .12834,13288.. .13360, 13958.. .14072,14674.. .14859,15291.. .15425)) /note="strong similarity to AMP-binding protein - Brassica napus; Contains Putative AMP-binding domain signature AA255-266; contains EST gb:A1999282.1" /codon_start=1 /product="AMP-binding protein" /protein_id="CA578449.1" /db_xref="GI:7268112" /translation="MASTSLGPTLLSYGSPSRQPDFGFLRISGHESVRIPSRFR REXKVPSPFLSSSFGDAALRSSEWKAVIDWRSAEKYGDVALVDPPYHDPPLK LTYKLEQETLDEAGELVGLGVKADSKIALFADNSRWLVSDQAVNVVRGSRSSVEE LQIYRHSESVAIVVONPEFFNRIAESFTSKASLRELLILWGEKSSLVTOGMOIPVYS YAEIINQGESRAKLSASNDTRSYRNQFIDSDTAAIMYTSGTGPKGVMLTHRNLL HQIKLSKYPVPAQAGKDFUSLPSWHAIERASEYFIFTCGEQMYTSIRYLKDDLKRY EQPNYIVSVLYETLSYGKQKQSASAGRKFLALILIKVSMAYMEMKRIYEGMCLTK EQPPMYIVAFVDMARVIAALLPLHLMLAKKLIYKKIHSSTIGISKAGISGGSLPI HVDKFEAIGVILQNGYGLTETSPVVCARTLSCNVLGSAGHPMHGTEFKIVDPETNNV LPPGSGILIKVGPQYMGVYKNPSTTKOVLNMSGWENTGDTGWIAPHHSKGRSRHCG GVVILGKAKDTIVLTGENVEPLEIEEAAAMRSRVIEQIVVIGQRRRLGAILIPNKE EAQRVDPSTKSLKSLVYQELRKWTSECSFQVGPVLIIVDDPTIDNGLMTPTMKIRR DMVYAKKEIDQLYS" complement(9655.. .9749) /number=1 complement(9750.. .9807) /number=2 complement(9808.. .9882) /number=2 complement(9883.. .9998) /number=3 complement(9999.. .10096) /number=3 complement(10097.. .10176) /number=4 complement(10177.. .10275) /number=4 complement(10276.. .10435) /number=5 complement(10436.. .10593) /number=5 complement(10594.. .10731) /number=6 complement(10732.. .11064) /number=6 complement(11065.. .11145) /number=7 complement(11146.. .11246) /number=7 complement(11247.. .11300) /number=8 complement(11301.. .11382) /number=8 complement(11383.. .11541) /number=9 complement(11542.. .11652) /number=9 complement(11653.. .11758) /number=10 complement(11759.. .11881) /number=10 complement(11882.. .11949) /number=11 complement(11950.. .12155) /number=11 complement(12156.. .12305)	intron	/number=13 complement(12555.. .12625)
		exon	/number=13 complement(12626.. .12834)
		intron	/number=14 complement(12835.. .13287)
		exon	/number=14 complement(13288.. .13360)
		intron	/number=15 complement(13361.. .13957)
		exon	/number=15 complement(13958.. .14072)
		intron	/number=16 complement(14073.. .14673)
		exon	/number=16 complement(14674.. .14859)
		intron	/number=17 complement(14860.. .15290)
		exon	/number=17 complement(15291.. .15425)
		exon	/number=18 complement(16137.. .16174)
		gene	16137.. .17732 /gene="AT4g14080"
		CDS	complement(join(16137.. .16174,16252.. .17604,17687.. .17732)) /note="the cDNA A26668 is cognate to this gene; gene A6 for anther-specific protein - beta-1, 3-glucanase homolog; Contains Glycosyl hydrolases family 17 signature AA274-287;Prokaryotic membrane lipoprotein lipid attachment site AA8-18; contains EST gb:AW004339.1, Z18400" /codon_start=1 /product="A6 anther-specific protein" /protein_id="CA878450.1" /db_xref="GI:7268113" /translation="MSLLAFFLTLVSSSCCSATRFQGRHYMQRTMLDLASKIGI NYGRGNLPSQYQINFIKIRAGHVLYDADPELTLSTNLYVTITVPHNQITA LSSNQTIADQWRTNIPYYPQIRVFLVNEILSYSGNVSVNLVPAMRKIVNSLR LHGITHNKVGTPLAMDSLRSFPSTNGTFRREITGVPVMLPLKFLNCTNSYFFLNVHP YFRSRNPMTSLDEALFOGSHYTDQTLGVYENLLDQMLDSVLFAMTKLGYPHML A1SETGPNFGDIDETGANTLNAATYNRNLIKMSASPPITGTPSRGCLPIPTFFVSLE NENKSGSGTORHWGIILHPDGSPIYDVDTGQTPLTGPNLPKPTNNVYKQVWCVP VEGANETELETIRMACAQNNTCAALAPGRECYEPVSIYWHASYALNSYWAQFRNQS IQCFNGLAHETTTPGNDRCFKPSTL" complement(16175.. .16251) /number=1 complement(16252.. .17604) /number=2 complement(17605.. .17686) /number=2 complement(17687.. .17732) /number=3 18189.. .18494 /gene="AT4g14090" /number=1 18189.. .25761 /gene="AT4g14090"
		intron	join(18189.. .18494,18694.. .18787,18938.. .19018,19571.. .19724, 19772.. .19833,21489.. .21524,22014.. .22792,22844.. .22948, 23060.. .23200,23280.. .23566,24375.. .24631,24686.. .25140, 25192.. .25311,25445.. .25614,25665.. .25761) /gene="AT4g14090"
		exon	/note="similarity to proton-coupled peptide transporter PEPT, Rattus norvegicus; Contains Prokaryotic membrane lipoprotein lipid attachment site AA535-545;Prokaryotic membrane lipoprotein lipid attachment site AA955-965; contains EST gb:A1998193.1"

Query Match	19.9%	Score	300	DB	8	Length	198944
Best Local Similarity	55.7%	Pred.	No. 2.5e-55				
Matches	768	Conservative	0	Mismatches	530	Indels	80
Gaps							
Qy	36	tgctagcaagcttcttcgcgaagccacataaatccgcctccattcgcaagagac	95				
Db	21201	TGCTGTGTAAACATCCCGAGCGAAGGTACATCAACCGCGCGCTTCAACTAGCCAAACCGCC	21142				
Qy	96	tcctaaaagcggcgaactgaagtcacatttttcacagagcggttatgcatggcgccgcatgg	155				
Db	21141	TCATCCACCAGGTCGAACCGTCACATACCTCCACCGGAGTCTCTGCTCACGCGATGG	21082				
Qy	156	ccaaacagcctccgcgcgtgcgcgaaacccacccgggocctgaattcgtgcgttcctccg	215				
Db	21081	CGGAGCCACCTTCCA-----CAAAAGGTCTATCCTCGCTGGTTTCCCG	21037				
Qy	216	acggctacgacgagggctgaagccctgcgcgagcgggaagcgctacatgctccgagatga	275				
Db	21036	ATGGATTTCGAGCAGCGGTCTCAAGTCATTTCGAAGACCAAGAAAATCTAGATGCCAACTCA	20977				
Qy	276	aagcccgcggtccgcgagccttaagaaacctctctcacaacacagacg-----	326				
Db	20976	AACGATGTGGTTCAACCGCCTGAGAGACATCATCAAGCCAACTTGTGACGCCACCACCG	20917				
Qy	327	-----tcaogttogtctactccccacctcttggatggcgcgaggtcgg	374				
Db	20916	AAACAGAGCCTATCACCGGGTAACTACTCTGTCTCTCGTCCCGTGGGTTTCTACGGTAG	20857				
Qy	375	cgctgaagtcacgggtcccgagcgcctctctgggtcgagccgcgcacccgtgctgca	434				
Db	20856	CGCGTGAGTTTCACCTCCCAACTACACTTCTTGATTTGAACACGAGTACTGTACTAGACA	20797				
Qy	435	tatatcttacttctcaacggctcagcagacgagatcagccggttcccgacgaattc	494				
Db	20796	TCCTACTACTACTTCAACACCTTTACAACATCTCTTCGAGGTTGAACCGATTAAAT	20737				
Qy	495	agctccctcggtctccaccctcgagcagcgagcttctccgacctttctgctgcgcgaga	554				
Db	20736	TACCGAAACTGCCACTGATCACCCGCGTACCCTCCCGTCTTTCTTCAACCTTCGAAAG	20677				
Qy	555	caccggagagatccggttgatgatgaaggagagagctggaaactttagacgggtgaagaga	614				
Db	20676	CATTACCGCTAGCTCTTTGTGACTCTAGAGAAACATATATGGAAGCTCTCGAAACGGGAATCA	20617				
Qy	615	aggcgaaagtgttggtgaacacgcttctgctggtggagccgagtcgactcacgctattg	674				
Db	20616	ACCTAAGATTCTTTGTTAAACATCTCTGCTTGTGGAAACACGATGCTTTAACCTCTGTTG	20557				
Qy	675	atagggtatgattgattcgggatcgggcgttgattccctccgcctcttggacgcgcgag	734				
Db	20556	AGAAACTCAAGATGATCCCAATCGACCGTGTGGTTTCTTCTCTCCGAGGGTAA-----	20503				
Qy	735	atccctcgaacgctcttacggcgcgcatctttcgaaaaatacgggagagataactgcg	794				
Db	20504	-----AACCAGATCTTTCAAATCTTCGACGAGGAT---TACA	20470				
Qy	795	tggagtgttgacacgaacgaatacttcggtggigtatgctgctgttgga-----gcg	851				
Db	20469	CGAATGTTTAGATCTGNAAGTCGAGAGATCAGTGTATTTACATTTCTTTAGGCACACACG	20410				
Qy	852	ttttgaggtttccaaagggcacaataatggaagagattgggaaggggctattagcttcgcgaa	911				
Db	20409	CCGATGATTTACCAGAGAAACACATGGAAGCGTTACTACGCGGTGTTAGCTACAAACA	20350				
Qy	912	ggcggttttttgatgatcagcagacagaagaatgacacgcgcgaagagaagaagaag	971				
Db	20349	GACCGTTTTTATGGATCGTGAAGGAGAAAAATCCAGAG-----AGAAAGAAGAAGAT	20297				
Qy	972	agtgatgttcattgggggaattgaaaaaaatgggga-aaatagtttctgctgtctcgcag	1030				
Db	20296	CGGTTCTTTGAATTTGATCAGAGGAAGTGATCGAGAGTTGGTGGGATGCTGTTCTCAG	20237				
Qy	1031	ttggaggtcttcggcgaccctcgcttgggattgttcgtgcagcattgtgggtggaactcg	1090				

Db	20236	ACAGCTGTTTGGCGCATTTGTGCTGGGATGTTTTTGACTCATTTGGTGGAAATTCG	20177
Qy	1091	gctgtgcagacttgagtgcgggttcaggtagtgcgggtgcggcgaatggtttgatcacg	1150
Db	20176	ACGTTGGAGAGTTTAGAGAGTGTGTCCGGTGGTTCCTTTCCGCAGTTTGCTGATCAG	20117
Qy	1151	acgacgaatcgaaactgattgaggatcgctggggacaggggtgagagtgagaataaat	1210
Db	20116	TGTACAACGCCGAAGCTTGTGGAGGATACGTGGAGGATTGGAGTCAAGTGAAGTTTGGG	20057
Qy	1211	gaaggggggttgaatgatctgagatgagagggctgctgagagatggtgatgattggg	1270
Db	20056	GAGGAAGCGAGATGTGGATGGGGAGGAGATTAGAAGCTGTTTGGAGAAGTGATGAGTGT	19997
Qy	1271	ggtgagaagcaaacatgtaagcaaaaaatgccataaaaatgaagacaccttggccagaaa	1330
Db	19996	GGAGAAGCGCGGAGGAGATGAGAGAGATGCAGAGAATGGAAGCGGATGCTTTGAT	19937
Qy	1331	gccatgggagaggtgatcttcactcaagaatctcaacgctttcttcacaaattg	1388
Db	19936	CGCGCACGGAAGGTGGACGTCGGATTGAATCTTAAGAGTTTGTGGACGAGGATG	19879
RESULT	6		
LOCUS	ATFCA0	200576 bp DNA	PLN 28-JUN-1999
DEFINITION	Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment	No. 0.	
ACCESSION	297335		
VERSION	297335.2	GI:5280985	
KEYWORDS	.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi		
AUTHORS	Bevan, M., Stiekema, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N., Kreis, M., Kavanagh, T., Entlan, K.D., Rieger, M., James, R., Puigdomenech, P., Hatzipoulos, P., Obermaier, B., Duesterhoft, A., Jones, J., Palme, K., Ansoorge, W., Delseny, M., Bancroft, I., Mewes, H.W., Schueeller, C. and Chalwatzis, N.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 200576)		
AUTHORS	EU Arabidopsis sequencing, project.		
TITLE	Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schuell@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7JU Norwich, UK, E-mail: michael.bevan@bbcsc.ac.uk		
COMMENT	On Jun 30, 1999 this sequence version replaced gi:2244747. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATFCAI at the 3' end.		
FEATURES	Location/Qualifiers		
source	1..200576 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4"		
source	1..63289 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /clone="BAC TAMU20K8" /size=.4428 /gene="dl33000w" Join(1753..3194,3261..4428) /gene="dl33000w"		
gene			
CDS			

```

/note="similarity to Cf-2.2, Solanum pimpinellifolium"
/codon_start=1
/product="disease resistance Cf-2 like protein"
/protein_id="CAB10171.1"
/db_xref="GI:2244748"
/translat="SPTREMBL:O23253"
/translat="MTI1WSLCIFLCSNLSILVIARDLCLPQDQDALLEFRNEFSTP
SPSDMLAIOTAKRNNTDCCSGGISCDDPKTVGVVVELDGLNGLRNSNLSIF
RYHDLQSLDLSYNDLCTLPDSSNGFYLRVLNLCNLFGEIPTSLSISYLDL
SLNDLGTGLDSMGNLKHULRVLSLSCHFTKIPSLGNLTFLTDLDSWNYFTGEL
PDSMGNLURVUNLHRCNFFKIPISLGLSNLTDIDISKNFTSEGDPSMSLNL
TDLQMLNLSLTYNDLSNQPKAMLPNMSLSLEAPDISGNFSFGTIPSLFML
PSLIKLDLTNDFSGPKTGNISPSNLQELYICENNINGPIPSILKLVGLSALS
FMTGIVFDSFLOKLSRSLDLSGLNLNLISSSHLPMMHLISSLNCSQFPKFL
ENOTSILYHLDISANOEGOVPEWLRLPTLSFIASDNKFSGETPRACETGTVLNN
NFSGSIIPCFEISNKVLSILHRLNLSLGVIPESLHGVLRSIDVGNRLSGQPKSL
INCSYLOFLNVEBRLNDFPWSLKLPLQLLVLNRSNEHGFIFSGDLSFSKURF
FDISENRFSYDPSDFVGNVMSFVIDNTPGTVVGDQDESFKHSVVLTKGLN
MELVSGFSEYKTDIVSGNRLEDGIDFESIGILKLVNKNNAFTGHIPPSLSNLSN
LQSLDLSQNLKSGSIPGELGELTFLARMNFSYNLLEGPDPQGTQIOSONSSSFAENPG
LCAPLQKCGGEEDKEKEKDKGLSWAAAGVYVPGFCGLAIGHILTSYKRWFE
MRIFSCFSPLKK"
exon
1753..3194
/gene="dl3000w"
/number=1
intron
3195..3260
/gene="dl3000w"
/number=1
exon
3261..4428
/gene="dl3000w"
/number=2
exon
complement(5905..6149)
/number=1
5905..7913
/gene="dl3000c"
complement(join(5905..6149,6243..6344,6436..6826,
7236..7913))
CDS
/note="strong similarity to glycine
hydroxymethyltransferase (EC 2.1.2.1); contains EST
gb:1585775"
/codon_start=1
/product="hydroxymethyltransferase"
/protein_id="CAB10172.1"
/db_xref="GI:2244749"
/db_xref="SPTREMBL:O23254"
/translat="MEPVSWGNTSLVSDPEIHDLEKRRQCRGIELIASENFTS
FAVIEALGSALTWKYSEGIPGNRYGNEFIDEIENLRSALEAFHCDPAAGVNVQ
PYSGSPANFAAYTALLOPHDRIMGLDLPGGHLYTHGYTSGGKKISATSYFESLPYK
VNFYTGIDYDLEEKALDPRKLLICGGSAYPRDWDYARFRAIDKVGALLCDMAH
ISGLVAAQEAANPEYCDVVTTHKSLRGPAGMIFYRKGPYPKKGQPEGAVYDFE
DKINFAVFPALQSGPHNHQIGALAVALKOANTPEKVKYAKQVKANAVAGLNYLSKGY
QIVYNGTENHLVLDLRLPLGLTGNKVEKLDLCSITLKNNAVFGDSALAPGGVYRGA
PMTSGRLVKEDEQIGEFLSRAVTLFDIQTYTGKLLKDFNGLVNNKDLQKADV
EKFSASYEMPGFLMSMKYKD"
intron
complement(6150..6242)
/number=1
exon
complement(6243..6344)
/number=2
intron
complement(6345..6435)
/number=2
exon
complement(6436..6826)
/number=3
intron
complement(6827..7235)
/number=3
exon
complement(7236..7913)
/number=4
gene
12823..14568
/gene="dl3010w"
join(12823..13533,13822..14568)
CDS
/note="strong similarity to adenosylhomocysteinase (EC
3.3.1.1)"
/codon_start=1
/product="adenosylhomocysteinase"
/protein_id="CAB10173.1"
/db_xref="GI:2244750"
/db_xref="SWISS-PROT:O23255"
/translat="MALLVETSSGREYKVKDMSQADFGRLELELAEVEMPGLMACRT
EFGSPQPFKARITGSLHMTIQAVLITETLALCAEVRWCSCNIFSTQDHAATAIARD
SAEVFAWKGTLEQYVWMTBRALDWGPGGPDLDIVDGGDATLLIHGVKAEIIFKFT
GOVDPDTSTNDPQFIVLSITIKELQVDPKPKYHKMKERLAVGVSETTTGVKRLVQMOQ
NGTLTPAINVDSVTKSKFDNLYGCHSLPDGLMRATDVMIAKGVAVICGYDVGKG
CAAMKTAGARVIVTEIDPICALQALMEGLQVLEDDVSEADIFVTTGNKDLIMVD
HMKMKNNANVCNIHGFNDIMLGLTETPGVKRITIKPQTDWRVFTTKAGIILVLA
GRMLNLCATGHPSEFVMSCSFTNQVLAQLNEKASGKYKKYVYVLPKHLDEKVAL
HLKLGARITKLSQDOSDYSIPIEGPKPHRY"
exon
12823..13533
/gene="dl3010w"
/number=1
intron
13534..13821
/gene="dl3010w"
/number=1
exon
13822..14568
/gene="dl3010w"
/number=2
exon
complement(16093..16270)
/number=1
16093..16501
/gene="dl3015c"
complement(join(16093..16270,16317..16501))
CDS
/note="weak similarity to Walleye dermal sarcoma virus gag
polyprotein"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB10174.1"
/db_xref="GI:2244751"
/db_xref="SPTREMBL:O23256"
/translat="MFNSTALVIFAILFLISADAFPIPSNGSIDAMLIRNSIIGED
DELMPTESIRRVLAQKRNQELLTIVALKLILTAEAATPLDALETQTTSTHEGKI
QETAKGSLAEASKLYD"
intron
complement(16271..16316)
/number=1
exon
complement(16317..16501)
/number=2
CDS
complement(join(18573..18914,18997..19137,19212..19554,
19603..19719,19762..20123))
/note="weak similarity to fusion protein la/lb - sugar
beet"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB10175.1"
/db_xref="GI:2244752"
/db_xref="SPTREMBL:O23257"
/translat="MDHVSSLPOEVLYHLSFTTKEAALTSLSKWRNLFTVPVNL
DIDDSVFLHQEKGEDRYETQKSFMKFVDVRVALQGNSPIKKLSKLRTGDFSHRVDG
WISNALARGVTELDLLILNLVTLKNSVRVDMLAAGDIFLPMUKTLVLHRSVRLCVDN
VWNGKDVVNSASLKILTYNICLTGFSVDTSVYFCFTEYVAIDYVYIMONLF
EARTSLVTDQTERAPNIDVDDVLRVNMKILKIRGVQVYSPNTLEVLA
SLCCSEMPVNNLKSLLTKSHSRGQWAMPVLLRNCPLFTLVLEGLLHVHTDKGVA
CDCVSRREKCSLTSCPVKLEIKGFRGTIKEMHIKHFVNYFPCLMETNIYIESGP
PKLRVQVCEFIAMHEHYKSSFNQVLLVGGYLYKKTAQ"
exon
complement(18573..18914)
/number=1
18573..20123
/gene="dl3020c"
complement(18915..18996)
/number=1
intron
complement(18997..19137)
/number=2
intron
complement(19138..19211)
/number=2
exon
complement(19212..19554)
/number=3
intron
complement(19555..19602)
/number=3
exon
complement(19603..19719)

```



/translation="MENLKNECHVLIALFPQGHINPLOSFKLLINLGKVKVLSSSL  
SAFNRIKNLEPKIEGLTPAPESDYGNGFKSGFDYHLFNSAIKSHSEFIANLIKSKA  
KNQPFTRVITILMDWAGSVAKKLHPSTLFWIOPATVIVYRTFNANFYKNYD  
SQDILLEPCGLPSLSSDPSFVEDDWSNDWAVESKKROIELINSEENPRILVNFED  
ALEBNALRVLKNVTMGVIGLPISSFLDEKRDKNFTFAADIMSEENYMWLDARANK  
SVIYIAPGSVAEISSOMMEIEISQGLKGRPFLWVINTLNGEKPEKUTCKDELEKI  
GRIVRMSQMEVLEKHSVGFLLHCHWNSTLESAGVPVACPIWINDQICNAKLIOQ  
VMKIGVNVANKEGII1IKRDEFQKCIETVMGDAEGBELRKNQAQWKDLAKESTKENS  
SNVNLKAYVNECLLGHG"

polya\_site

/note="35 A nucleotides"

BASE COUNT 514 a 238 c 325 g 517 t

ORIGIN

Query Match	18.88;	Score 283;	DB 7;	Length 1594;
Best Local Similarity	60.99;	Pred. No. 1.1e-51;		
Matches 501;	Conservative 0;	Mismatches 310;	Indels 12;	Gaps 2;
Qy	578	atgaaggagaaagctggaaactttagacggtgaagagagcgaaagtgttgtaaacacg	637	
Db	640	ATTAAAGGCAAAATGAGATATTGAATAGTCAGAAATCCAGAAATCTTGTGAATACT	699	
Qy	638	tttgatcggttgagccgatgcactcacggtctattgatagtgatgattgattcgggatc	697	
Db	700	TTTGATGCATTGGAACCTTAATGCTTTGAGGGCTTTTGAAAAATGTAACATGTTGGTAATT	759	
Qy	698	gggcggttgattccctccgctctttgagacgcgagagatccctccgaaacgtcttaagcc	757	
Db	760	GGACCTTTAAATTCCTTCATCTTTCTTGATGAGAAAGATCGAAAGATPAATTTTTTCGCG	819	
Qy	758	ggcgtatctttcgaataacgagagagaaataactgcgtggagtggttgacacgaagccg	817	
Db	820	GCTGATATGATCGAAA-----GTGAAATAATTACATGGAGTGTTCGATCGCAAGGCG	873	
Qy	818	aaatcttcggtggtgtatgtctgttgggagcgcttttggaggtttccaaagcgcaaatg	877	
Db	874	AATAAATCAGTATTATTACATAGCATTTGTTAGCTATGCTGAGATATCAAGTCAATGGATG	933	
Qy	878	gaagagattggaaaggctattagctcgcggaaggccgtttttatgatatcagaaa	937	
Db	934	GAAGAGATTAGTCAAGGATTGTTGAAATGTGGAAGACCATTTTGTGGGTGATTAGAGAA	993	
Qy	938	cagaagaatgcgacgcggaagaaagaagaagattgagttgattcggttggaattgaaa	997	
Db	994	ACCCTAAAT-----GGGAAAACCCGAGGAAAGTTGACTTGTAAAGATGAATTGGAA	1047	
Qy	998	aaatggggaaaaatgatttcgtggtgcgcagttgaggttctggcgcaacctgcgttg	1057	
Db	1048	AAGATAGGAGAAATAGTGCCTTGTGTCGCAAAATGGAAGTCCCTAAAACACTCGTCCGTA	1107	
Qy	1058	ggatgtttcgtgacgcatgtgggtggaactcgctgctggagagcttgattgcggggtt	1117	
Db	1108	GGCTGTCTTCTTGACTATTGTTGGATGGAATTCGACTCTCGAGACCTTAGCTTGTGGTGTG	1167	
Qy	1118	ccggtggtggtgcgtggtgcgcagttggtgatcagacgacgaatgcgaagctgatgagat	1177	
Db	1168	CAATCGTGGCATGCCCCGATTGGAAATGTAATTTGCAACGCGCAAGCTCATTCAGAT	1227	
Qy	1178	gcgtgggggacaggggttgagatgagaatgaatgaagggtggtggttgatggatctgag	1237	
Db	1228	GTTTGGAAAGATTGGTGTGAGAGTGAACCGCTAATAAGAAAGCTATCATTTAAAGAGACCGAG	1287	
Qy	1238	atagagaggtgtgtgagagatggtgatggtggtggtgagagagacaaactagttagagaa	1297	
Db	1288	TTCCAGAGTGTATAGAGATCGTAATGGAGATGCTCGAGGAGGAGAGGAATTCGAGAAAG	1347	
Qy	1298	aatccataaataagagactttgcccagagaagccatggagaggtgaggtatccttcaactc	1357	
Db	1348	AATGCCCAAAATGGAAGGATTGGCTAAGGAATCTACGAAGGAAATAGTTCATCTAAT	1407	
Qy	1358	aagaatctcaacgctcttctcctcaatcaagttcagctgctgcttaat	1400	

Db	1408	GTAAATCTTAAGGCTTATGTTAAATGAATGTTTACTTGTGTCATT	1450	
RESULT	8			
AC005106/c				
LOCUS	AC005106	84203 bp	DNA	PLN
DEFINITION	Genomic sequence for Arabidopsis thaliana BAC T25N20 from Chromosome 1, complete sequence.			
ACCESSION	AC005106			16-DEC-1999
VERSION	AC005106.2	GI:5587719		
KEYWORDS	HTC.			
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; brassicaceae; Arabidopsis.			
AUTHORS	Shinn, P., Buehler, E., Dewar, K., Feng, J., Kim, C., Li, X., Sun, H., Conway, A., Conway, A., Kurtz, D., Oji, O., Shen, Y. K., Toriumi, M., Vysotskaia, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and Ecker, J. R.			
TITLE	Genomic sequence for Arabidopsis thaliana BAC T25N20 from Chromosome 1			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 84203)			
AUTHORS	Ecker, J. R.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA			
REFERENCE	3 (bases 1 to 84203)			
AUTHORS	Ecker, J. R.			
TITLE	Direct Submission			
JOURNAL	Submitted (31-JUL-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA			
REFERENCE	4 (bases 1 to 84203)			
AUTHORS	Ecker, J. R.			
TITLE	Direct Submission			
JOURNAL	Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA			
REFERENCE	5 (bases 1 to 84203)			
AUTHORS	Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Conn, S., Kim, C., Altafi, H., Bei, B., Chio, C., Chio, J., Choi, E., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaver, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A. and Ecker, J.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-DEC-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA			
COMMENT	On Dec 16, 1999 this sequence version replaced gi:3366536.			
FEATURES	Location/Qualifiers			
source	1. .84203			
	/organism="Arabidopsis thaliana"			
	/db_xref="taxon:3702"			
	/chromosome="1"			
	complement(join(1491..2159,2389..2452,2568..2703,3061..3149,3231..3277,3369..3457,4022..4139,4246..4353))			
	/note="5' end of hypothetical protein YUP8H12.1; overlaps T7 end of YUP8H12"			
	/codon_start=1			
	/evidence=not_experimental			
	/product="T25N20.1"			
	/protein_id="AAC80580.1"			
	/db_xref="GI:3935137"			
	/translation="MGKQKDHSHVSVESDDKVEAVLHLKKHSPSLTKQKFCNRRACVGRLRIKGDNVKKAQKQLRSLSRSLGIESLTADEFTAEALGAYVAGLDECRPVLVRIKQDYQKLHTQKLTQLRLVLTLEVAISTMSRNVQEVILEDAIAAALLVTL			

CDS	/db_xref="taxon:3702"			
	/chromosome="1"			
	complement(join(1491..2159,2389..2452,2568..2703,3061..3149,3231..3277,3369..3457,4022..4139,4246..4353))			
	/note="5' end of hypothetical protein YUP8H12.1; overlaps T7 end of YUP8H12"			
	/codon_start=1			
	/evidence=not_experimental			
	/product="T25N20.1"			
	/protein_id="AAC80580.1"			
	/db_xref="GI:3935137"			
	/translation="MGKQKDHSHVSVESDDKVEAVLHLKKHSPSLTKQKFCNRRACVGRLRIKGDNVKKAQKQLRSLSRSLGIESLTADEFTAEALGAYVAGLDECRPVLVRIKQDYQKLHTQKLTQLRLVLTLEVAISTMSRNVQEVILEDAIAAALLVTL			



Db 69820 GTGTCCTCTCTCTCATTCACCCGCTCTATCATCCCAAAACCACAAAGCTCGCAAAAT----- 69764  
Qy 190 gggcctcgacttcgtggtctctccgacggtacgacgacgagcggtgtaagccctcgcggcga 249  
Db 69765 ----CTCTCTTTCTCTTACTTTCTCGAGGATTCGACGAGGATCTCAACACCGCA 69710  
Qy 250 cgggaagcgctacatgtccg--agatgaagccgcggtcccgagcgcttaagaacct 306  
Db 69709 CGAGCTCAAAACCGGTGGTACACTTCGAACGTAATGGCCGTAAGCTCTATCGGATTT 69650  
Qy 307 ccttctcaacacccacga-----cgtcacgttctgctctactctccaccctctt 354  
Db 69649 CATGAAGCTAATCAGATGCTGACTCTCCCGTAAGTTGCTGTGATCTACACGATCTTCC 69590  
Qy 355 tgeatggcgcgagaggtggcggtgagtcaccaggtcccgagcgccctctctgggtcga 414  
Db 69589 CAACCTGGGTTCCAAAGGCGCGGTAGATTTCATCTTCGCCCTGTTTCATCTCGATCCA 69530  
Qy 415 gcccgccacgctgctgtgcataatattacttctacttcaacggtacgcgacgagatcga 474  
Db 69529 ACCAGCCTTCGCTTCGCACATTTATTACAATTAATCTCTACAGG-----AAACAA 69482  
Qy 475 cgcggttccgacgaataatcagctccctcgcgtctccaccctggagcgcagtccttcc 534  
Db 69481 CTCCGTTTTCAGTTCCCGAATCTACCTTCTCTCGAATCCGCGATCTGC-----CTTCT 69427  
Qy 535 gaccttctgctgcgggagacacgcgagagatccggttgatgataaggagagactgga 594  
Db 69426 TTCCTCTCACCTCCACACGCAACAAACCCGCAACGAGTATATCAA-GAATGATGGA 69368  
Qy 595 aactttagcgtgaaggaagcggaagtgttgtaacacggttgatcgcttggaacc 654  
Db 69367 TTTTCTCAAAAGAAGTAATACCCGAAATTTCTCGTCAACACATTCGATTCGCTGGAGCC 69308  
Qy 655 cgatgcactcagcggcttagtgatgagttgatcgggatcgggcggttgattccctc 714  
Db 69307 AGAGTTCTTAACAGCTATTCCGAATATAGAAATGGTGGCAGTGTGCTCTTACTTCTCTGC 69248  
Qy 715 cgccttcttgacggcgagatccctccgaaacgcttcctacgcggcgatcttcttogaata 774  
Db 69247 GGAGATTTTCACTCGAAGCGAATCAGGTAAAGATTTATCAAGAGATCATC----- 69196  
Qy 775 atcggagagataactcgtggagtggttgacacgaagccgaatacttcggtggtata 834  
Db 69197 -----AAGTAGTAGTTATACACTTTGTTGTAGCTCGAAGAACAGAGTCTCTGTATTATTA 69143  
Qy 835 tgtgtcgtttgggagcgtttgagtttccaaaggcacaataatggaagagattgggaagg' 894  
Db 69142 TGTCTTTTGGACAATGGTTGAGTTGTCGAAGAAACAGATAGAGGAACATAGCGAGAGC 69083  
Qy 895 gctattagcctcggaagcgcggtttttatgatgatacagaaacagaagaatgacgacgg 954  
Db 69082 ACTCATAGAAGGGGAAGACCGCTTCTTGGGTGTATACTGATATAACTATAAATCAACAGAGAAGC 69023  
Qy 955 cgaagaagaagaagagattgagttgcatiggg-----gaatt 993  
Db 69022 GAATAAGAAGGAAGAAGAGACAGAGATTGAGAAGATAGCTGTTTATAGACAGAGCT 68963  
Qy 994 gaaaaaaatgggaaaaatagttcgtggtcgtcagttgaggttcttggcgccaccttcg 1053  
Db 68962 TGAAGAGCTTCGGATGATTGCTCGTGTGTCGAGATAGAGGTTTGTGACACCGAGC 68903  
Qy 1054 gttgggatagttctgacgcattgtgggtgaaactcggctgtgagagacttgagttgcgg 1113  
Db 68902 CATAGGTTGTTTTTGTACTCATTTGTTGGTGGAGCTCATCTGGAGAGATTTGGTTCTCGG 68843  
Qy 1114 gttccggttggtgcgggtgcgagtggtttgatcagacgcaatgcgaagctgattga 1173  
Db 68842 CGTTCCAGTGTGTGGCTTCCGATGTGGTTCGATCAGCCAGCAATGCGAAGCTTTTGGGA 68783  
Qy 1174 ggatgcgtgggggacaggggtgagagtgagatgaatgaagggggttgggttgatgc 1233

Db 68782 AGAAATATGAAGACACAGCTGTGAGGCTGAGAGAGAACTCGGAAGCTTTAGTAGAGAGAGG 68723  
Qy 1234 tgaatagagaggggtgtgtgagagatggtgatggatggggtgagaagagcaaaactagtga 1293  
Db 68722 AGAGATAATCGGTGTTTGGAAAGCAGTGTATGGA-----GGCGAAATCGGTGAGCTGAG 68669  
Qy 1294 agaaaatgccataaaatggaagactttggccagagagacatggagagagatggatcttc 1353  
Db 68668 GGAAACCCAGAGAAATGGAGCGTTTAGCGACTGAAGGGGTAGAGAGGAGGATCTTC 68609  
Qy 1354 actcaagaatctcaacgcctttct 1377  
Db 68608 GGACAAGAATGTGGAAGCTTTGT 68585  
  
RESULT 9  
LOCUS AC007153 103223 bp DNA PLN 17-MAY-1999  
DEFINITION Arabidopsis thaliana chromosome I BAC F3F20 genomic sequence,  
complete sequence.  
ACCESSION AC007153  
VERSION AC007153.2 GI:4580365  
KEYWORDS HTG.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
Arabidopsis.  
REFERENCE 1 (bases 1 to 103223)  
AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
Alfafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,  
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,  
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.S.,  
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.  
Unpublished  
JOURNAL 2 (bases 1 to 103223)  
REFERENCE Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
Alfafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,  
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,  
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,  
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.  
Direct Submission  
JOURNAL Submitted (24-MAR-1999) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
3 (bases 1 to 103223)  
REFERENCE Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
Alfafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,  
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,  
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,  
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.  
Direct Submission  
JOURNAL Submitted (08-APR-1999) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
4 (bases 1 to 103223)  
REFERENCE Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
Alfafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,  
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,  
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,  
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.  
Direct Submission  
JOURNAL Submitted (17-MAY-1999) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
COMMENT On Apr 9, 1999 this sequence version replaced gi:4508084.  
Bases 1-33654 of clone of clone F3F20 overlap with bases  
50550-84203 of 'RAMU' clone Y25N20, gb|AC005106.  
e-mail for correspondence: arabsequence.stanford.edu  
Genes with similarity to proteins in the databases are described as  
'putative', '-like' or 'similar to'. Genes that have EST  
similarity but no significant protein similarity are described as

'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge, <http://genomic.stanford.edu/~chris/GENSCANW.html>), Fexa (V. Solovyev & A. Salamov, Sanger Centre, <http://genomic.sanger.ac.uk/>), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/NetPlantGene.html>).

## FEATURES

```

source
    1..103223
        /organism="Arabidopsis thaliana"
        /cultivar="Columbia"
        /db_xref="taxon:3702"
        /chromosome="1"
        /clone="F3F20"

misc_feature
    1..33654
        /note="overlap with bases 50550-84203 of 'TAMU' clone T25N20, gb|AC005106. Features from bases 1 to 29,000 are annotated on overlapping clone T25N20, gb|AC005106."
        complement(29404..40201)
        /gene="F3F20.1"
        complement(join(29404..29652,29751..29936,30023..30073,30150..30950,31055..31284,31428..31486,31568..31651,31740..31858,31981..32100,32252..32411,32506..32695,32780..32848,32961..33086,33172..33281,33357..33508,33634..33700,33851..33948,34038..34161,34243..34403,34491..34575,35347..35441,35543..35680,35806..35892,35975..36044,36112..36220,36644..36981,37155..37272,37354..37473,37725..37818,37915..37991,38079..38204,38319..38445,38591..38682,38787..38879,38964..39080,39165..39245,39382..39505,39608..39729,39823..39923,40015..40201))
        /gene="F3F20.1"
        /note="Highly similar to putative callose synthase catalytic subunit; Highly similar to cotton putative callose synthase catalytic subunit, gb|AAD25952."

gene
    /codon_start=1
    /protein_id="AAD30609.1"
    /db_xref="GI:4836907"

CDS
    /translation="MAQRPPDPDPQRRILRTQVSGISGEMLSDEVVPSSVETAPILRVANEASPRVAYLCRYFAFEKAHLDPDTSRGVRQFKTALLQRLERENETTLACRQKSDREMOSFYQHYKRYIQALLNAADRAQLTKAYQTAALFVELKAVNQTEDEVAHGKKLDEDILDKIQIYVNIPLDPDSONAIMRLPEIQAAVAALNRTRGLPWTAGHKKLDEDILDLQSMFGQKQNVLNQREHLILLANVHIROPKPDQOQPKLDRATLVNKKLFYKWKCKYLGKSSMLPTIQQEVQKRLLYGLLYLWGEANLRPMCELCIYTHHLYGMLAGSVSPMTGEHVKPYGGEDEALQKVVYPIYQTSKEAKRSRGSKSHVWRYDDLNIEYFWSIRCFPLGMPMRADAFECOTAEELRLERSEIKNSGDRMGKVNFEIRSEWHIFRSDRLWSFYILCQAMIVIAWNGSGELSAIFQGDVFLKLSVFTITAILKLAQVALDIALSWAKHSMYLVKRYVMKGAAGAAVWVVMVAVTAYSNKASGFSOTKNWFGHSHNSPLFIVAILLYLSPNLSALLFLFPFRRLYLESDYKTMMLMWNQSOPSPSSQIKPLVPTKDIRJHSIYWSHEFPFHAKNNLGVVIALMSPVLYTFMDQIWAIVLYSTVLGNGAFRGLRIGIRLGMRSRQSPKIPGAFNDCVLVDNDDDTKMKFRATFRSKFDQLPSKDKKAARQAMWNKIISFRREDLSDREMLLWYSDPDLDLWPPFLIASKIPALOMAKDNGKRLKRLKALVDSTMTCAVREYASFKENLYLVVGEREGOVINDIFSKIDHIEKETLITLESALPDLVGOFYRLLEYLNEEDKQDQIVLVLANMLELVTRDIMEEVEYPSANISWNFDOSQFLIKKLGRKKQIKRLHLLLVKESAMDVPSNLEARRLITFSNLSLMDPAPKLRNMLSPSVLYTFSDDVLSIFGLTEQNEGDVSLFYLQKIFPDWTNFTLVRKVCNGEELRAREDLHEELRNLWVVEIRSGOTTKTVRGVMYIRKALEQALFDMARDELLKGYKALELTSSEASKSGSLWACQALADMKFTFVVSQCOQSYHKRSQDQAKDILRLMTYPSIRVAYIDEVQTHKESYKTEEKIYYSALVKAAPOTKPMDSSESVOITDQIYRIKLPAGIILGEGVQTHKESYKTEEKIYYSALVKAAPOTKPMDSSESVOITDQIYRIKLPAGIILGEGKPEONHAIYTFRGEGLQITDMNQNYMEEAFKMRNLQOEFLEKKGVRCPITLIGREHFTGSVSLAFNMFMSNQSFSVYIGORVLASPLKVRPHYGHDPIDFRLHLRFGFNSLZREGNVTHHEYIOVGRGDRVGLNQISWFEAKIANGNEQTLRDLRLGLRGLDFPNMLSCYFTIGTFYFTMLTVLYVFLVGLRVLVSLGEBLSQRFAFNKPLPEAAASQSFVQIGFMAIPMMMEIGLERGFHNAIEFLVQLQLASVFFTFQIGTKTHYIGTRFLHGAAYRGTRGCVFVFAFNAENRYFSRSHVKGIELMILLVYQIFGQSYRGVVTYILITVSYFMVVTWLPAPFLNPSGFMQKIVDDMTDNKWIYNGGIGVPRKSWESWKEKEHLRHSHSGRGITLIEFLARFFIFQYGLVYHLSTFKGKQNSFWYGASWFIWFLILVLVGLGVRGRFRFTNFQILFRIRIKGLVFLTFVAILITFLALPLITIKDLFCMILAPMPTGCMGLLIAQACKRPLIQOLGIWSSVRLTARGYEIVMGLLLFTPTVAFLAWFFPY

```

Query Match

14.8%; Score 222.8; DB 8; Length 103223;



[illegible]

Db	18413	TGAAGAGGTGGGATGATGTCTCGTGTGTTCGAGATAGAGGTTTGGAGACACCGAGC	18354
QY	1054	gttgaggatgttcgtgacgcatttggtggaactcggtctggagagcttgagttgcgg	1113
Db	18353	CATAGGTGTTTTTGTACTCATTGTGGGTGAGCTCATCTGGAGAGTTTGGTTCGCG	18294
QY	1114	ggtccggtggtggtgacgcatttggtggaactcggtctggagagcttgagttgcgg	1173
Db	18293	CGTTCAGTGTGTGCGTTCGATGTGTGCGATCATGCCAGCAATGCGAAGCTTTTGG	18234
QY	1174	ggatcggtggtggtgacgcatttggtggaactcggtctggagagcttgagttgcgg	1233
Db	18233	AGAAATATGGAAGACAGAGTGTGAGGGTGTGAGAGCAACTCGGAAGCTTTAGT	18174
QY	1234	tgatagagaggtggtgacgcatttggtggaactcggtctggagagcttgagttgcgg	1293
Db	18173	AGAGTAATGCGGTGTTGGAGCAGTGTGGA-----GCCGAAATCGGTGGAGCTG	18120
QY	1294	agaaatgccaataaaatggaagactttggccagagaagccatggagaggtgagttc	1353
Db	18119	GGAAACGACAGAAATGGAAGCGTTTAGCGACTGAAGCGGTGAGAGAGGAGGATCTTC	18060
QY	1354	actcaagaatctcaagcctttct	1377
Db	18059	GGACAAGAAATGTGAAGCTTTTGT	18036
RESULT	10		
ATCHRIV42/c			
LOCUS	ATCHRIV42	195068 bp	DNA
DEFINITION	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 42.		PLN
ACCESSION	AL161542		
VERSION	AL161542.2	GI:7268303	
KEYWORDS	thale cress.		
SOURCE	Arabidopsis thaliana		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1. (bases 1 to 195068)		
AUTHORS	EU Arabidopsis sequencing, project.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevand@bsrc.ac.uk		
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a> this fragment has an overlap with ATCHRIV41 at the 5' end and an overlap with ATCHRIV43 at the 3' end.		
FEATURES	Location/Qualifiers		
source	1. 195068		
	/organism="Arabidopsis thaliana"		
	/variety="Columbia"		
	/db_xref="taxon:3702"		
exon	/chromosome="4"		
	9917..10066		
	/gene="AT4g15560"		
	/number=1		
CDS	join(9917..10066,10350..10475,10559..10874,10981..11255,11340..11613,11691..11998,12081..12362,12441..12725,12813..12953)		
	/note="AT4g15560"		
	/gene="GT/AC sites ignored to preserve continuity; light inducible gene required for chloroplast development; Contains Prokaryotic membrane lipoprotein lipid attachment site AA452-462; Prokaryotic membrane lipoprotein lipid attachment site AA591-601; Transketolase signatures AA102-121; Transketolase signatures AA505-521; contains ESTs EST1		

	gb:AW004332.1, AI997953.1, W43562, N96015, W43388																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
--	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

[illegible]









numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, and T13116, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jining Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: att@tigr.org.

#### FEATURES

```

source      Location/Qualifiers
            1..99188
             /organism="Arabidopsis thaliana"
             /cultivar="Columbia"
             /db_xref="taxon:3702"
             /chromosome="II"
misc_feature <1..1802
             /note="Sequence from clone T9H9"
             complement(<1064..>2005)
             /gene="At2g31720"
             complement(<1064..>2005)
             /gene="At2g31720"
             /note="T9H9.24; predicted by genscan"
             complement(<1064..>2005)
             /gene="At2g31720"
             /note="hypothetical protein"
             /codon_start=1
             /db_xref="GI:6598601"
             /translation="WVRGDLNGLGGKRALDILCVATRTVSFAYEIGETSTEG
             DKRRERLRRLINSLSTMTNSSLKRLKIDCKNQDPEQNPVAVSPSSCHLES
             KRQKVSNNKPVNRKPVITDEDRPTTMLIDVMEVGMADAKLIFVRLPNSDYDE
             LQTRLMPKQILDMFLNEELEKIDRHKKISASDGKADIVVNSKGLQRLKLR
             WDMTSTNVVLGWNKVTNIDLQRTLRIRWSPDMLFFAFVLSDDPDPAPTECL
             NLLAKICEETTCLEALQEAQRNSSLVSDTELELR"
             complement(1803..>99188)
             /note="Sequence from clone F20M17"
             complement(Join(3588..3973,4058..4094,4189..4265))
             /gene="At2g31730"
             complement(3588..4265)
             /gene="At2g31730"
             /note="F20M17.23"
             complement(Join(3813..3973,4058..4094,4189..4260))
             /gene="At2g31730"
             /note="unknown protein"
             /codon_start=1
             /db_xref="GI:4887763"
             /translation="WGTDAVLLGEMQYIQFLQGVKCSLSASNPFFYRVEYSILRS
             KGLCLPEYTSVAGTADIAWPKYKPTTSSHAFNLSSNSPQ"
             complement(Join(<7083..7286,7407..7557,7647..7813,
             7921..7995,8070..8150,8243..8367,8449..8698,9242..9289,
             9386..9469,9576..9833,9973..10145,10486..10555,
             10653..>11039))
             /gene="At2g31740"
             complement(<7083..>11039)
             /gene="At2g31740"
             /note="F20M17.22; predicted by genscan"
             complement(Join(7083..7286,7407..7557,7647..7813,
             7921..7995,8070..8150,8243..8367,8449..8698,9242..9289,
             9386..9469,9576..9833,9973..10145,10486..10555,
             10653..11039))
             /gene="At2g31740"
             /note="hypothetical protein"
             /codon_start=1
             /protein_id="AAD32298.1"
             /db_xref="GI:4887762"
             /translation="MGKKKGKNAASTDDFLQTLDTSKENWDKFFTLRGNDSPFW

```

```

YAEWPQLHSLPLFLQDSSSSSDSLQILVPGCGNSRLTEHLVDAGFRDITNVDFSKV
VSDMLRIRTRPELRWRMVDITKMLADESFDVLTKGALDADPEVALFLSRFR
FKMKMNVHSIAOKRSLKTYMVAEKENSVLLEHTSAFELVSLGENDSGQSGMREAV
ESENKTRRCNNGSDLLYSHEDLKLGKGLAELETRIKFTLGGQGNFSYRAVLL
DAQKQEPYVYCGVFLVPKTRAHEWLFCEGQWQVVESSQAARLIMFLDSSHSGA
TMEDQVTPMTGKVVEDVYVESAPSNLEDLSLSDAFRLVFKRTEGLIQSEALL
VDEGTLEQSKQETKVNQSKRKNKQKQSPSRPLMRVSHDYLAASSYHTGIISGFT
LVSSYLKKAESCCTMVKTVVIGLAGLLPMFLHGLCPFFSIEAVELDPPVMLSVGKYF
GFTQNDLRVHIADGKIFKIRDTNSEASSESNIGSDSTHNTQGGICPCDILLID
VBSADSSGGLTSPASDIETELLVSVKALPQHGLFIVNLVTRRSQYKDMVVRMKRY
FDHLFGLQLEEDDVNVLFGLCSESVISENDIPESAVILEGLLKQCRLETKQSIIIDA
TKKLCWK"
             10152..10235
             /rpt_family="(TAA)n"
             complement(12661..12688)
             /rpt_family="(TAAA)n"
             join(13281..13986,15175..15987)
             /gene="At2g31750"
             13281..15987
             /gene="At2g31750"
             /note="F20M17.21"
             join(13342..13986,15175..15900)
             /gene="At2g31750"
             /codon_start=1
             /product="putative glucosyltransferase"
             /protein_id="AAD32297.1"
             /db_xref="GI:4887761"
             /translation="MGEKAKANLVSPFIQGHINPLLOFSKRLLSKNVNTLTTSS
             THNSILRRATGGATPLPSFVIDDGFEDHPSTDTSPDYFAKGVNSRSLSEILIS
             SMDKPNNAVYDSCPLVDYCKRKHGVAASFTQSTVNATYIHLRGEPEFQND
             VVLPAMPPLKGNLDPVFLVDNNLCRFLFELISSQFVNDDIDFLVNSFDELEVLQ
             WKNQWPKNIGMIPSMYDLDRKLGDKYGINLFNAQVNECLDLSKPPGSVIVVS
             FGSVLVKDDQMLEVAAGLKQTCGHNPLWVRETETKLPNSYIEDICDKGLIWNMSPQ
             LQVLAHSGCCFWTHCGWSTLEALSIGVALICMPAYSDPTNAKIEDYWKVGVRYK
             AQNGFVPKEIYKRCVWEDMSKGRKRNKRLMEFAREALSDGGNSDKNIDEF
             VAKIV"
             22837..22997
             /rpt_family="(CAAAA)n"
             23089..23130
             /rpt_family="POLY_A"
             join(<24523..25698,25968..26061,26146..>26420)
             /gene="At2g31760"
             <24523..>26420
             /gene="At2g31760"
             /note="F20M17.20"
             join(24523..25698,25968..26061,26146..26420)
             /gene="At2g31760"
             /codon_start=1
             /product="putative ARI-like RING zinc finger protein"
             /protein_id="AAD32296.1"
             /db_xref="GI:4887760"
             /translation="MDYSDDMIDNESGNNSDGGNENYNNAAVDITLSEKSYV
             IKEELIKQRODIERTILFSQVAIVLILHLMVSKLEDEFTDEERIKTV
             GILKEPVDLVNTEVDIQGCFEYSYTRKEIASVCGHPYCKTGWYITFKLEDGP
             CLVRKPEPCSYAVQGMIDETKDKDYRYFLRSYVEDGKMKMKNPSCGCECA
             VEGESSGDVACLQSYRCWNCSDAHSVDCTVSKYKIFKQDSEKNMILANSK
             PCPKRPLEKSHGCHNMTCSASCHRFWCCKSYSDHVACNNYVEDADHDKRTLLQ
             SEKRTHTYVVRWENQSSRLKAMSDLEKFSQVLQKLSQDCKPKIDQLQFVDANLQ
             IIECRRLKWTYAYGYLLDNLRKPLFEYLQGEAETGLERLKAENELKQFFIKSED
             PSDTFNAFRMKLTGLTKVTNTPDNLVKALENGLADVTKSSESDAFLETKLYDIYI
             SEGCF"
             27052..27180
             /rpt_family="(TAAA)n"
             join(<27609..28814,29071..29164,29249..>29580)
             /gene="At2g31770"
             <27609..>29580
             /gene="At2g31770"
             /note="F20M17.19"
             join(27609..28814,29071..29164,29249..29580)
             /gene="At2g31770"
             /codon_start=1
             /product="putative ARI-like RING zinc finger protein"
             /protein_id="AAD32295.1"

```







Db 241 CTGCGGCGACGGGAAGCGGCTACATGTCCGAGATGAAGCCCGCGGCTCCGAGGCCTTAAG 300







```

Qy 581 ----aagagagagctggaacatttagacggtgaagagaaagcgaaagtgttggtgaacac 636
Dy 931 ATTCAGAAGAAATGATAGACACATTAGATGCTGAACACCACTCTAGATCTTGTGGAATAC 990
Qy 637 gttgatcggttgagcccgatgcactcaocggctattgatagtgatgattgacgagat 696
Dy 991 ATTTGATGAATTAGAGCCTGAGGCACCTCAATGCAATTAAGGTTATAAGTTTATGGAAT 1050
Qy 697 cggccggtgattccctccgctctcttgacgcgagcgagatccctccgaaacgctttacgg 756
Dy 1051 TGGACCGGTGATCTCTGCTTCTTGGGTGGAATGACCCCTTAGATGCTTCATTGCG 1110
Qy 757 cggcgatcttttcgaaataacgagagaaataactcggtgagtggttggtgacagaaagcc 816
Dy 1111 TGGTGATCTTTTCAAAATCAATGACTATA-----TGAATGGTTAACTCAAGGCC 1164
Qy 817 gaaatctcgggtgtatgtgtcgttgggagcggttttgaggtttccaaaggccacaaat 876
Dy 1165 AAATTCATCAGTGTGTTATATATCTTTTGGGAGTCTAATGAATCCATCTATTAGCCAAAT 1224
Qy 877 ggaagagattgggaaaggtattagctcggaagcgccgtttttatggtgatgacagaga 936
Dy 1225 GGAGGAGATATCAAAAGGGTTGATAGACATAGGAAGGCCGTTTTTATGGGTGATAAAGA 1284
Qy 937 acagaagaatgacgacgcggaagaagaagaagagtggtgattgcttggtggaattgaa 996
Dy 1285 AAATGMAAAAGCCAA--AGAAAGAGAGANTAAAGAGCTTGGTGTATTGAAGAAATGGA 1341
Qy 997 aaaaatggggaaaaatagtttcgtggtcgctgcagttgaggtttctgcgcacctgggtt 1056
Dy 1342 AAAAAATAGAAAAATAGTTCATGCTGTTCACTCAACTTGAAGCTTCTAAAAATCCATCTTT 1401
Qy 1057 gggatgttcctgacgcaatttggtggaactcggtcgtggagagcttgattcgggggt 1116
Dy 1402 AGGATGTTTTGCTCTCAATTGGAATGGAATTCAGCCCTTAGAGAGTTTATGCTTGGAGT 1461
Qy 1117 tccggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1176
Dy 1462 GCCAGTGTGCAATTCCTCAATGACAGATCAAAATGACAAATGCCAAACAAAGTTGAAGA 1521
Qy 1177 tgcgtgggggacaggggtgagagtgagatgaatgaagggggtggtggtggtggtggtggt 1236
Dy 1522 TGTGTGGAAGAGTGGAGTAAAGAGTGAAGTAAATGAAGATGTTGTTGAAAGTGAGGA 1581
Qy 1237 gatagagaggtggtgagagatggtgagatggtggtggtggtggtggtggtggtggtggt 1296
Dy 1582 AATCAAAAGGTGTATTGAATTTGGTAATGGATGGAGGAGAGAGGAGGAGAAATTTGAGAAA 1641
Qy 1297 aatgccaataaattggaagactttggccagagaaagccatggagaggtggtggttcaact 1356
Dy 1642 GAATGCTAAGAAATGGAAGAAATTTGGCTAGAGAGAGTGTGAAGGAGTGTGATCTTACA 1701
Qy 1357 caagaatctcaacgccttcttcacaaagtgc 1389
Dy 1702 CAAGAAATTAAGAGCTTTTATTGATGATGTTGC 1734

```

## RESULT 5

```

X02829 ID X02829 standard; cDNA; 1671 BP.
AC X02829;
DT 14-MAY-1999 (first entry)
DE W09905287 Seq ID 4.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Torenia hybrida.
FH Location/Qualifiers
FT 45..1481
FT CD5 /*tag= a
FT /product= "protein with flavonoid 5-transglycosylation
FT activity"
FT W09905287-A1.
PN 04-FEB-1999.
PD

```

```

PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR ) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI; 99-142940/12.
DR P-PSDB; W92950.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
used to transform plants for improvement of plant coloration
PS Disclosure; Page 60-64; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC the production of flower varieties with desired coloration and cut
CC flowers harvested from them.
SQ Sequence 1671 BP; 530 A; 344 C; 399 G; 396 T;

```

```

Query Match 28.7%; Score 433.2; DB 1; Length 1671;
Best Local Similarity 62.5%; Pred. No. 8.5e-113;
Matches 889; Conservative 2; Mismatches 435; Indels 96; Gaps 10;

```

```

Qy 26 cgcgcgtgctgtagcaacgtttcctcgcaagcgccacataatcccgccccaattc 85
Dy 57 CGCCATATTTCTACTAGCAACATTCACGACCAAGGCCACATAAACCCTTCTCGAGTTC 116
Qy 86 gccaagagactcctaaagccgg-----cactgacgtcacatttttcacagcggttat 139
Dy 117 GCCAAAAGGGCTCCACACACCGGATACGTCGACCAAGTCACATTTCTCAGAGTGATAC 176
Qy 140 qcatggcgcccatggcccaacacacagctcccgctgcgggaaacccaccggccctgac 199
Dy 177 GCATTGACGCATGCGGCTTCGAAACCGATCGGACGACAGAA-----TCGAT 224
Qy 200 ttctggcggtttccgacggtctacgacgacggtggaagccctcgcggaagcgagcgc 259
Dy 225 TTCGTGGCATKTYCAGATTCTTACGATGATGGCTTAAAGAAAGGCGACGATGGCAAAAAC 284
Qy 260 taactgctcagagatgaagcccgcggtcccgagccttaagaaacctctctcacaac 319
Dy 285 TACATGTCGGAGATGAGAAGCGCGGAACGAAGGCGCTTAAAGGACACTCTTATTAAAGTC 344
Qy 320 caccga-----cgtcaggtctcgtctactcaccac 349
Dy 345 AACGATGTCGATGGGAAGTGAATGTTACAATCCCGTGAGCTTTGGTGTGACTCTCAT 404
Qy 350 ctctttgcatggcgcgaggtggtggtggtggtggtggtggtggtggtggtggtggtggt 409
Dy 405 CTATTTTCTGGCAGCTGAAGTGGCGGTGAAGTCGACGTGCCGAGTGCCGCTTCTTTGG 464
Qy 410 gtcgagcccgccaccgtgctgcatatattactctacttcaacggtacgacagag 469
Dy 465 ATTGAACCGGTACGGTTTTTCGATGTGTACTATTTTACTTCAATGGGTATGCCGATGAT 524
Qy 470 atcgacgcggttccgacgaaattcagctccctcggtctccaccctcgagagcgagcgt 529
Dy 525 ATCGATGCGGCTCAGATCAAAATCCCACTGCCCAATCTTCCGACGCTCTCCAAAGCAAGAT 584
Qy 530 ctccgacctttctgctcgcgagacacccgagagattccgggttgatgatgaagagagaag 589
Dy 585 CTCCCTCTTTCTCTACTCTCTTCGAGCCCGCGAGATCCGAACCCCTAATGAAGAAAAAG 644
Qy 590 ctggaaccttttagacggtgaagagaaagcggtggtggtggtggtggtggtggtggtggt 649
Dy 645 TTCGACACGCTCGCAAAAGAACCCGAAGCGAAGGTCTTGATAAACAACACTTCGACGATTA 704
Qy 650 gagcccgatgcatcaccggtattgataggtgatggtggtggtggtggtggtggtggtggt 709
Dy 705 GAACCGAACAACTCAAAAGCCATCGACAGGTATGAACATAATATCCATCGGCCCATTAATC 764
Qy 710 cccctccgc---cttcttgacggtggtggtggtggtggtggtggtggtggtggtggtggt 760
Dy 765 CCATCATCATATTTCTCAGATGGCAACGACCCCTCATCAAGCAACAAATCTACGGTGA 824

```







QY	624	tgttggtgaacacgcttttgatgctgttggagcccgatgcactcagcgctattgataggtatg	683
Db	703	tgtttcggtcaacacgcttcgaggggttagagacacaaaggtttgtaagtttgagaaatcagagg	762
QY	684	agttgatcg-----ggatcggccgcttgattccctcgccttcttggacgcggagatc	737
Db	763	cgatgagggcgagcttgatcgagacctatgataccatctgcttatctcgacggccgaatca	822
QY	738	ctctccgaacgcgtctacggcggcgactcttttcgaaaaatcggaggagaaataactcgtgg	797
Db	823	aagacgataaagcctacggctcgagccctgat- --gaaagccgctctcgaggagtgctatgg	879
QY	798	agtggttggacacgaagccgaaatcttcggtgggtatgtgttcgttttggggagcgttttga	857
Db	880	agtggttagacactaagctgagcaagtcggtggttttttgggttgcgtttggttcccttggga	939
QY	858	ggtttccaaaggcacaaatggaagagattgggaaggctattagcctgcggaagccgct	917
Db	940	tcctctttgagaagcaactcgcctgagctggcacaaggcgtttacacgaatccaacttttaact	999
QY	918	ttttatggatgatacgaacacagagaatacgaacgcgcgaagcgaagaagaagagtgta	977
Db	1000	tcgttgggtgatcaaaagacgtctatagcgaagttaccgaag-----	1044
QY	978	gittgcatgggggaattgaaaaaatgggaaaaatagtttcgtgctcgcagttggagg	1037
Db	1045	ggtttgtggaagctaccaaaagacagagcgttcgttcttgggtgtatccagcgttgagg	1104
QY	1038	tcttgccgcacccctgcgttgggatgttctcgtgaacgattgtgggtgaaactcgcctgttg	1097
Db	1105	tttttagctcatgtatccatagggttgcgtttttgactcactgcgggttgcaactcgacgttg	1164
QY	1098	agagcttgagttgcggggttcccggtgggtgcgcagctgttgcgttgatcaagacacga	1157
Db	1165	aagatattcagcttttgggagttcccgatgggtgtgtgccttcagtggagtgatcagatgaatg	1224
QY	1158	atgcaagactgattgaggatgcgctgggggacaggggtgagagtgagaatgaatgaagggg	1217
Db	1225	atgcttaagtttggtagagaggtttgagagattgggtatagaccgaaggagaaagcttgggg	1284
QY	1218	gtggggttgatggaatctgagatagagaagtgctgtgagagtgttgagatggtatgagtgag	1277
Db	1285	caggagttgtgaagaccgatgaggtgtgaggtgtttgacaggagtgatggaaggagaga	1344
QY	1278	agagcaactagtgagagaaaaatgcataaaatggaagactttggccagagaagccaatg	1337
Db	1345	gtagtgtggagatttagagagagttctaagaagtgsaaagatttggctgtggaagccgatga	1404
QY	1338	gagagatggatcttcaactcaagaatctcaacgccttct	1377
Db	1405	gtgaaggaggaagctctgatcggagacatttaactcagatttgg	1444

## RESULT 7

Tl0106	
ID	Tl0106 standard; cDNA to mRNA; 1731 BP.
AC	Tl0106;
DT	27-OCT-1996 (first entry)
DE	Maize UDP-glucose indol-3-yl-acetylglucosyl transferase sequence.
DE	UDP-glucose-indol-3-yl-acetylglucosyltransferase; enzyme;
KW	transgenic plant; EC-2.4.1.121; crop improvement; corn; cereal;
KW	grass; IAGlu-transferase; ds.
OS	Zea mays.
Key	
FH	Location/Qualifiers
FT	57..1472
FT	/tag= a
FT	858
FT	/tag= b
FT	misc_binding
FT	/note= "start of UDP binding region"
PN	WO9600291-A1.
PD	04-JAN-1996.
PF	20-JUN-1995; U07820.

24-JUN-1994; US-265427.  
(RESE) RESEARCH CORP TECHNOLOGIES INC.  
PA Bandurski RS, Szczygłowski K, Szerszen JB;  
WPI; 96-068875/07.  
DR P-PSDB; R95999.  
DR UDP-glucose indol-3-yl:acetyl:glucosyl transferase DNA and protein  
PT - used in sense or anti-sense orientation for the genetic control of  
PT plant tAA levels and plant growth  
PS Claim 2; Page 21-23; 41pp; English.  
CC This nucleotide sequence encodes a UDP-glucose-indol-3-  
CC ylatetyl:glucosyltransferase (IAGlu-transferase) from maize (Zea mays).  
CC Over-expression of this sequence in transgenic plants will result  
CC in at least 1 of the following effects: inhibition of apical  
CC dominance, inhibition of stem elongation; inhibition of cell  
CC enlargement and increased numbers of stems per plant as compared  
CC with a wild-type plant. The ability to control the expression of  
CC this enzyme allows the control of free indoleacetic acid levels in  
CC plants, thereby effecting plant growth rates. Antisense constructs  
CC may be used to inhibit synthesis of this enzyme.  
CC Sequence 1731 BP: 298 A; 528 A; 601 G; 304 T;  
SQ

Query Match 8.7%; Score 131.8; DB 1; Length 1731;  
Best Local Similarity 53.7%; Pred. No. 1.4e-27;  
Matches 324; Conservative 0; Mismatches 267; Indels 12; Gaps

[illegible]

RESULT  
T66173  
8



```

FH Key      Location/Qualifiers
FT CDS      2..1410
FT          /*tag= a
FT          /product= "glucosyl transferase"
FT          /note= "Glucose from amino acid position 5
FT          shown in W47172"
PN WO9745546-A1.
PD 04-DEC-1997.
PF 30-MAY-1997: G01473.
PR 31-MAY-1996: GB-011420.
PA (UYTO-) UNIV YORK.
PI Bowles DJ, Calvert CM, Odonnell PJ, Roberts MR;
DR WPI; 98-032653/03.
DR P-PSDB; W47172.
PT Tomato wound inducible (TWI1) gene encoding glucosyl transferase -
PT useful to develop products that alter signalling pathways in plants
PT by altering of salicylic acid, jasmonic acid or ethylene
PS Claim 1; Fig 1; 52pp; English.
CC This wound inducible gene (TWI1) isolated from wounded tomatoes encodes
CC a glucosyl transferase (GTase) protein. This TWI1 gene can be used to
CC identify homologue GTase encoding genes isolated from tobacco and rice.
CC A microbial host can be transfected or transformed with a vector
CC containing the GTase encoding nucleic acids. The products can be used to
CC interfere with GTase and therefore alter signalling pathways in plants,
CC specifically tobacco, rice or tomato plants by altering levels of
CC salicylic acid, jasmonic acid or ethylene. This can induce the
CC production of plant defence proteins such as pathogenesis-related (PR)
CC and proteinase inhibitor (PIN) proteins which regulate plant development
CC (plant growth, reproduction and senescence) and improve plant response to
CC pathogens.
SQ Sequence 1624 BP; 545 A; 252 C; 372 G; 455 T;

Query Watch      5.0%; Score 76; DB 1; Length 1624;
Best Local Similarity 50.5%; Pred. No. 8.3e-12;
Matches 221; Conservative 0; Mismatches 205; Indels 12; Gaps 1;

QY 779 gagagaataaactcgctggagtggttgacacgaagcgaatcttcggtggtgtatgtg 838
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 779 GATGACACCGCTGCTTGAATGGCTTGATTCGAGAAATCAAGTCCATGTTTATGTT 838

QY 839 tcgttgggagcgcttttgaggtttccaaaggcacaatggaagagattgggaaggccta 898
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 839 TGTGTTTGAAGTACAGCAGATTTTCACTACAGCAGATGCAAGAACTTGTATGGGCTA 898

QY 899 ttgacctgcggaagcgcttttcatgatgatcagacagaagaatgacgacgcgaa 958
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 899 GAAGCCTCTGGACAAGATTTTCAATTTGGTTATCAGAACAGGGAAATGAAGATTGGCTCCCA 958

QY 959 gaagaagaagaagtggtgagttgcattggggaattgaaaaaaatgggaaaaatagttcg 1018
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 959 GAAGGATTCGAGGAAGAAC-----AAAGAAAGAAAGTTTAAATCATAGAGGA 1006

QY 1019 tgggtctgcagttgaggttctgcgcacccctgcgttggatggttctgcgacgattgt 1078
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1007 TGGGCACCCCAAGTGTGATCTTGATCAGGAAGCTATTGGAGCTTTTGTACTCATGTT 1066

QY 1079 ggtggaactcggctgtggaagactgagttgcgggggttcgggtggtgcgcgag 1138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1067 GGATGGAACTCGACACTGGAAGAAATATCAGCAGGGGTACCAATGGTGACATGCCAGTA 1126

QY 1139 tgggttgatcagacgacgaatcgagagctgattgagagtcgctggggacaggggtgaga 1198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1127 TTTGCCGACAGTTTTTCAATGAGAGTGGTGACTGAGTAATGAGACTGGAGCTGGT 1186

QY 1199 gtgagaatgaatgaaggg 1216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1187 GTTGGTCTTAAGCAATGG 1204

RESULT 11
X24873
ID X24873.standard; cDNA; 1626 BP.
```

```

AC X24873;
DT 21-JUN-1999 (first entry)
DE Medium-chain UDP-glucose:fatty acid glucosyltransferase cDNA.
KW UDP-glucose:fatty acid glucosyltransferase; tomato;
KW 1-O-acyl-a-glucose; acyl donor; esterification; ss.
OS Lycopersicon pennellii.
OS Lycopersicon esculentum.
FH Key      Location/Qualifiers
FT CDS      1..1416
FT          /*tag= a
PN WO9909144-A1.
PD 25-FEB-1999.
PF 10-AUG-1998: IB1369.
PR 30-JUN-1998: US-106464.
PR 13-AUG-1997: US-055554.
PA (CORR ) CORNELL RES FOUND INC.
PI Eanneatta NT, Ghangas GS, Kuai J, Steffens JC;
DR WPI; 99-190155/16.
DR P-PSDB; W98009.
PT New chain length specific uridine diphosphate-glucose:fatty acid
PT glucosyltransferases - useful for preparing straight or branched,
PT (unsaturated 1-O-acyl-a-glucoses and 1-O-acyl-a-D-glucoses
PT Example 3; Page 33-36; 42pp; English.
PS The present sequence encodes a UDP-glucose:fatty acid
CC glucosyltransferase that transfers glucose to medium chain length
CC fatty acids to form 1-O-acyl-a-glucoses. The 1-O-acyl-a-glucoses
CC act as acyl donors in the esterification of glucose and in the
CC further esterification of partially acylated glucose in reactions
CC catalysed by glucose acyltransferases. They also act as acyl donors
CC in the esterification of sucrose and in the further esterification
CC of partially acylated sucrose in reactions catalysed by sucrose
CC acyltransferases. The source of the present cDNA is leaf trichome
CC mRNA of an F1 population of a cross between wild tomato
CC Lycopersicon pennellii (LA 716) and cultivated tomato
CC Lycopersicon esculentum cv. New Yorker. The clone was isolated
CC following immunoscreening of a trichome cDNA library. The
CC invention provides chain length-specific UDP-glucose:fatty acid
CC glucosyltransferases and a method for their purification. The
CC enzymes have specificity for transferring glucose to short, medium
CC or long chain length fatty acids. Methods for preparing
CC 1-O-acyl-a-D-glucoses, where the acyl group comprises 3-5, 6-13
CC or 14-22 C atoms, are claimed.
SQ Sequence 1626 BP; 540 A; 256 C; 372 G; 458 T;

Query Watch      5.0%; Score 76; DB 1; Length 1626;
Best Local Similarity 50.5%; Pred. No. 8.3e-12;
Matches 221; Conservative 0; Mismatches 205; Indels 12; Gaps 1;

QY 779 gagagaataaactcgctggagtggttgacacgaagcgaatcttcggtggtgtatgtg 838
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 793 GATGACACCGCTGCTTGAATGGCTTGATTCGAGAAATCAAGTCCATGTTTATGTT 852

QY 839 tcgttgggagcgcttttgaggtttccaaaggcacaatggaagagattgggaaggccta 898
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 853 TGTGTTTGAAGTACAGCAGATTTTCACTACAGCAGATGCAAGAACTTGTATGGGCTA 912

QY 899 ttgacctgcggaagcgcttttcatgatgatcagacagaagaatgacgacgcgaa 958
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 913 GAAGCCTCTGGACAAGATTTTCAATTTGGTTATCAGAACAGGGAAATGAAGATTGGCTCCCA 972

QY 959 gaagaagaagaagtggtgagttgcattggggaattgaaaaaaatgggaaaaatagttcg 1018
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 973 GAAGGATTCGAGGAAGAAC-----AAAGAAAGAAAGTTTAAATCATAGAGGA 1020

QY 1019 tgggtctgcagttgaggttctgcgcacccctgcgttggatggttctgcgacgattgt 1078
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1021 TGGGCACCCCAAGTGTGATCTTGATCAGGAAGCTATTGGAGCTTTTGTACTCATGTT 1080

QY 1079 ggtggaactcggctgtggaagactgagttgcgggggttcgggtggtgcgcgag 1138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1081 GGATGGAACTCGACACTGGAAGAAATATCAGCAGGGGTACCAATGTTGACATGCCAGTA 1140
```

```

Qy 1139 tggttgatcagacgaagaatgcgaagtgtgattgagtgctggtgggacaggggtgaga 1198
      |||||
Db 1141 TTTCCGAACAGTTTTTCAATGAGAAGTGTGTGACTGAGGTAATGAGAGCTGGAGCTGGT 1200
      |||||
Qy 1199 gtgagaatgaatgaagg 1216
      |||||
Db 1201 GTTGGTTCTAAGCAATGG 1218
      |||||

RESULT 12
T68693
ID T68693 standard; cDNA; 432 BP.
AC T68693;
DE Strawberry UDP-glucuronosyl transferase cDNA.
KW Strawberry; UDP-glucuronosyl transferase; ERlib; fruit; ripening;
OS Storage; gene regulation; ss.
KW Fragaria ananassa Duch. cv. Brighton.
PN WO9721816-A1.
PD 19-JUN-1997.
PF 12-DEC-1996; G03076.
PR 13-DEC-1995; GB-025459.
PA (ZENE) ZENECA LTD.
PI Manning K;
DR WPI; 97-332787/30.
PT Vectors encoding new enzymes for regulating ripening of fruit - used
PT particularly to improve storage properties, processing
PT characteristics etc, especially in strawberries
PS Claim 1; Page 23-24; 43pp; English.
CC This cDNA sequence comprises a UDP-glucuronosyl transferase (ERT
CC lb) cDNA clone isolated from a ripe strawberry fruit cDNA library
CC by differential screening. 9 Ripening-enhanced clones (see T68687-
CC 95) clones were isolated from the library, and are deposited at
CC NCIMB 40693. Any of these 9 sequences can be used as a regulatory
CC sequence in claimed vectors for genetic transformation of plant
CC cells to regulate fruit ripening. The vectors also comprise a
CC promoter and transcription termination sequence. Also claimed are
CC plants (especially strawberry) and propagating materials containing
CC these vectors. Stable integration of the regulatory sequences, or
CC their complements or genomic equivalents, into a plant is used to
CC modulate fruit ripening by overexpression or downregulation of an
CC endogenous plant gene. Slowing ripening of fruit improves
CC resistance to damage during harvesting, packaging and transport.
CC It also extends shelf life, improves storage properties, processing
CC characteristics, flavour and aroma, and modifies colour and
CC increases resistance to post-harvest fungi.
SQ Sequence 432 BP; 101 A; 84 C; 136 G; 94 T;

Query Match 4.8%; Score 72.6; DB 1; Length 432;
Best Local Similarity 59.4%; Pred. No. 3.9e-11;
Matches 123; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 997 aaaaatgggggaaaatagtttcgtggtctgcagttggaggttctgcgcaccctgcgtt 1056
      |||||
Db 105 AGACAGAGCGCAAACTGTGCAATGGAGTCCACAGAGAAGATTTTGGAGCAGCTCTCGAC 164
      |||||
Qy 1057 gggatgttcgtgcagcattgtgggtggaactcgcgtggtgagagcttgagttcggggt 1116
      |||||
Db 165 GGCTTGCTTTGTGACTCATTCGGGTGGAACTCAACCATGAGTCACTCACTCAGGAAT 224
      |||||
Qy 1117 tccggtggtgcgvtgcgcagtggtttgtatcagacgaacgaatcggaagctgattgaga 1176
      |||||
Db 225 GCCCGTGGTGCAATCCCAACAAATGGGTGACCAAGTGCACCGCCAGCAAGTATTTGGTCGA 284
      |||||
Qy 1177 tgcgtgggggacaggggtgagatgag 1203
      |||||
Db 285 CGAGTTTAAGCTGGGAGTAACAATGTG 311
      |||||

RESULT 13
V23108
ID V23108 standard; cDNA to mRNA; 1597 BP.

```

```

AC V23108;
DE 06-AUG-1998 (first entry)
DE UDP-glucose:flavonoid 3,5-O-glucosyl transferase gene.
KW UDP-glucose:flavonoid 3,5-O-glucosyl transferase; gentian; rose plant;
KW blue flower; ds.
OS Unidentified.
FH Key Location/Qualifiers
FT CDS 39..1460
FT /*tag= a
FT J10113184-A.
PN 06-MAY-1998.
PD 15-OCT-1996; 272253.
PR 15-OCT-1996; JP-272253.
PA (TOLG) TOKYO GAS CO LTD.
DR WPI; 98-315473/28.
DR P-PSDB; W56451.
PT Gentian flavonoid glucosyl transferase gene - used to produce
PT recombinant rose plants with blue petals(s)
PS Claim 3; Pages 7-8; 10pp; Japanese.
CC The present sequence encodes a UDP-glucose:flavonoid 3,5-O-glucosyl
CC transferase enzyme. The nucleic acid sequence was amplified from
CC RNA which had been isolated from gentian (sic) petals. The gene is
CC used to transform rose plants to produce blue flowers.
SQ Sequence 1597 BP; 483 A; 272 C; 368 G; 474 T;

Query Match 4.6%; Score 69.4; DB 1; Length 1597;
Best Local Similarity 57.7%; Pred. No. 6e-10;
Matches 124; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 998 aaaaatgggggaaaatagtttcgtggtctgcagttggaggttctgcgcaccctgcgtt 1057
      |||||
Db 1035 AACCCGGAATGGTGAATGGATGGGCCCCACAGAGCGCGTGTCTCTCACTACAGCTCG 1094
      |||||
Qy 1058 gggatgttcgtgcagcattgtgggtggaactcgcgtggtgagagcttgagttgcggggtt 1117
      |||||
Db 1095 GGAGGTTTTCGTGTCGCAATTTGGATGGAATTCACACTTGAAGCAATTTGTTGGTGTG 1154
      |||||
Qy 1118 ccggtggtggcgggtgcgcagtggtttgatcagacgaacgaatcggaagctgattgagat 1177
      |||||
Db 1155 CCAGTGGCAACTGGCCCTCTGTTGCTGAGCAGCAGATGAATGCAATTTGAATTTGTTGAAG 1214
      |||||
Qy 1178 gctggggggacaggggtgagagtgagatgagaatgata 1212
      |||||
Db 1215 GAATGGGCTTGGCTGTGGAGGTTAAGATGGATTA 1249
      |||||

RESULT 14
T62123
ID T62123 standard; cDNA to mRNA; 1578 BP.
AC T62123;
DE 10-JUN-1997 (first entry)
DE Solanum melongena flavanoid-3-glucosyl transferase cDNA.
KW Eggplant; flavanoid-3-glucosyl; transferase; glucose; flavonoid;
KW pigment; production; seed; plant; ds.
OS Solanum melongena.
FH Key Location/Qualifiers
FT CDS 13..1314
FT /*tag= a
FT /product= flavanoid-3-glucosyl-transferase
FT J09056385-A.
PN 04-MAR-1997.
PD 25-AUG-1995; 238943.
PR 25-AUG-1995; JP-238943.
PA (KIRI) KIRIN BREWERY KK.
DR WPI; 97-206631/19.
DR P-PSDB; W13406.
PT Solanum melongena flavanoid-3-glucosyl transferase and DNA encoding
PT it - catalyses transfer of glucose to 3-position of flavonoid
PT pigment in plants
PS Claim 2; Page 10; 14pp; Japanese.
CC The present sequence encodes Solanum melongena (eggplant)
CC flavanoid-3-glucosyl transferase (F3CT), which catalyses the

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2000, 21:39:11 ; Search time 152.01 Seconds  
(without alignments)  
1363.704 Million cell updates/sec

Title: US-09-147-955-1

Perfect score: 1507  
Sequence: 1 gaaaattccacaaaatgg.....ttataatttagtaacaaaa 1507

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*

7: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131.8	8.7	1731	4	US-08-466-583-1
2	131.8	8.7	1731	6	PCT-US95-07820-1
3	84.2	5.6	7218	1	US-08-232-463-14
4	76	5.0	1627	5	US-09-106-464-1
5	66.8	4.4	1669	3	US-08-522-421-1
6	56.2	3.7	1607	4	US-08-797-226-1
7	55.2	3.7	405	4	US-08-975-316-32
8	54.4	3.6	1738	3	US-08-379-482A-2
9	46.6	3.1	28958	1	US-08-258-261B-6
10	46.6	3.1	28958	1	US-08-456-837-6
11	46.6	3.1	28958	1	US-08-457-342-6
12	46.6	3.1	28958	1	US-08-457-646A-6
13	46.6	3.1	28958	2	US-08-458-076A-6
14	46.6	3.1	28958	2	US-08-764-233A-4
15	46.6	3.1	28958	2	US-08-457-335A-6
16	46.6	3.1	28958	2	US-08-729-214-6
17	46.6	3.1	49377	2	US-08-764-233A-1
18	45.2	3.0	13987	3	US-08-804-227C-13
19	45.2	3.0	44377	3	US-08-804-227C-7
20	45.2	3.0	44377	4	US-08-804-198-1
21	44.6	3.0	1505	1	US-07-915-246-1
22	44.6	3.0	2336	6	PCT-US92-00282-1
23	44.6	3.0	2339	6	PCT-US92-00282-2
24	43.6	2.9	1358	2	US-08-471-033-45
25	43.6	2.9	1358	3	US-08-471-044-45
26	43.6	2.9	1358	3	US-08-463-483A-45

27	43.6	2.9	1358	3	US-08-471-046A-45	Sequence 45, Appl
28	43.6	2.9	1358	3	US-08-470-566B-45	Sequence 45, Appl
29	43.6	2.9	1358	4	US-08-469-334-45	Sequence 45, Appl
30	43.6	2.9	1358	5	US-09-300-529-45	Sequence 45, Appl
31	43.4	2.9	1338	3	US-08-484-993B-7	Sequence 7, Appl
32	43.4	2.9	1338	4	US-08-484-158B-7	Sequence 7, Appl
33	43.4	2.9	1338	4	US-08-484-596A-7	Sequence 7, Appl
34	43.4	2.9	1338	4	US-08-480-150A-7	Sequence 7, Appl
35	43.4	2.9	1338	5	US-08-458-731-7	Sequence 7, Appl
36	43.4	2.9	1338	5	US-08-149-223A-7	Sequence 7, Appl
37	42.8	2.8	2064	1	US-08-343-428-1	Sequence 1, Appl
38	42.2	2.8	2589	4	US-08-482-728A-3	Sequence 3, Appl
39	41.6	2.8	4897	7	5196516-7	Patent No. 5196516
40	41.6	2.8	53526	5	US-08-658-136-2	Sequence 2, Appl
41	41.6	2.8	53577	5	US-08-658-136-1	Sequence 1, Appl
42	41.4	2.7	434	3	US-08-332-766A-10	Sequence 10, Appl
43	41	2.7	450	4	US-08-387-942C-40	Sequence 40, Appl
44	41	2.7	12588	4	US-08-387-942C-1	Sequence 1, Appl
45	40	2.7	9960	5	US-08-822-586-46	Sequence 46, Appl

#### ALIGNMENTS

RESULT 1  
US-08-466-583-1  
; Sequence 1, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szareszen, Jędrzej B.  
; APPLICANT: Szczegłowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: and Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,583  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1731 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 57..1472  
US-08-466-583-1

Query Match	8.7%	Score 131.8	DB 4	Length 1731
Best Local Similarity	53.7%	Pred. No. 8.5e-27		
Matches 324	Conservative 0	Mismatches 267	Indels 12	Gaps 2
QY 781	ggagaataactgcgtgagtggttggacacgaagccaaatcttcggtggtggtatgtctc	840		
Db				
848	GGAGGATGCGTGACCAAGTGGGTAGACACAAAGCCGACCGCTCCGTGGCGTACGTCTC	907		
QY 841	gtttggagcgttttgaggtttccaaaggacaaaatggaagagattgggaagaagcctatt	900		
Db				
908	CTTCGGCAGGCTCCGTCCTGGCAACGCCACGAAGAGGAGTTCGCGCGCGCCTCCT	967		
QY 901	agcctgcggaagccgtttttatggtatgacagacagaagaatgacgcgcgcgaaga	960		
Db				
968	CGCCGCGGCAAGCCGTTCCTGTGGTGGT-----GAGGGCCACGACGACGACCA	1018		
QY 961	agaagagaagagttgattgcttcattggggaattgaaaaaatgggaaaaatagtttcgt	1020		
Db				
1019	GGTCCCGCGTATCTCCTGCCGAGGCGACGGCGACGGCGCCGATGGTCGTGCCCTG	1078		
QY 1021	gtactgcagattggaggtcttcggcacccctgcgttgggatgtttcgtgcgcattgtcg	1080		
Db				
1079	GTCCCGCAGCTGGACGTGTGGCGCACCCGCGCGTGGGTCTCTGCTACCCACTGGCG	1138		
QY 1081	gtggaactcgcgtgtgagagcttgattgcggggttccggtggtgcgcggtgcgcagtg	1140		
Db				
1139	TTGGAACTCCACGTGAGGCGCTACGTTCCGCGTGCCTATGTTGGCGATGGCGCTGTG	1198		
QY 1141	gtttgatcagacgacgaatgcgaagctgattgagtagtgcgtggggagacaggggtgagagt	1200		
Db				
1199	GACGGACACCGACCAACGCTCGGAACGTGACGCTCGCCTGGGCGCGGGCGCTCGCGC	1258		
QY 1201	gagaatgaatgaagg---gggtggggttgatgtagctgagatagagaggtgtgtggagat	1257		
Db				
1259	GCGCCGCGATGCTGGCGGCGCTGTTCCTTCGCGGGGAAGTGGAGCGGTGCGTGC	1318		
QY 1258	ggtgatgtagtggggttagaagagcaaaactagttagagaaaaatgccataaatggaagac	1317		
Db				
1319	CGTCATGGACGGGGGCGAGCGCGCTGTGCTGCACGAAGCGCGGGGGGAATGGAGGGA	1378		
QY 1318	tttggccagagaagccatggagagagtggtatcttcaactcaagaatctcaacgccttct	1377		
Db				
1379	CAGGGCTCGCGCGGTGGCACCCCGTGGCAGCTCTGACCGCAACCTGGACGAGTTCGT	1438		
QY 1378	tca 1380			
Db				
1439	GCA 1441			
Db				

```

RESULT      2
PCT-US95-07820-1
; Sequence 1, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1731 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 57..1472
; PCT-US95-07820-1
;
; Query Match 8.7%; Score 131.8; DB 6; Length 1731;
; Best Local Similarity 53.7%; Pred. No. 8.5e-27;
; Matches 324; Conservative 0; Mismatches 267; Indels 12; Gaps 2;
;
QY 781 ggaataaactcgtgagtggttgacacgaagccgaaatcttcggtggtgatgtgc 840
DB 848 GGAGGATGCGTGACCAAGTGGCTAGACACAGCCGACCGCTCCGTGGCCTACGCTC 907
QY 841 gtttggaagcgtttgaggtttccaaaggcacaaatggaagagattgggaaagggtcatt 900
DB 908 CTTTCGGCAGCCTCGCTGCCGTCCCTGGCAACGCCAGAGGAGGAGTCTCGCGCGGCTTCCT 967
QY 901 agcctcggaagccgcttttatggatgatacagagaacagaagaataagacagcgcgaga 960
DB 968 CGCGCGCGGCAAGCCGTTCTGTGGTGGT-----GAGGGCCAGCAGCAGCAGCACCA 1018
QY 961 agaagaagaagattgagttgcatggggaattgaaaaaaatggggaaaaatagtttctgt 1020
DB 1019 GGTCCCGCGCTATCTCTCGCCGAGCGCAGCGCGCGCGCGATGGTCTGTCGCCCTG 1078
QY 1021 gtgctcgcagttggaagttctgycacacctgcgttgggagatttctgtgacgcattgtg 1080
DB 1079 GTGCCCGCAGCTGACGTGCTGCGCACCCGCGCTGGGCTGCTTCGTACCCACATGCGG 1138
QY 1081 gtggaactcggctgtggaagacttgagttgcggggttccggtggtgagcggtgcgcgaagt 1140
DB 1139 TTGGAACTCCACGCTGGAGCGCTCAGCTTCGCGCTGCCATGGTGGCGATGGCGCTGTG 1198
QY 1141 gtttgcacagacaaatcggaagctgattggagatcgctgggggacaggggtgagagt 1200
DB 1199 GACGGACACCCACCAACGCTCGGAACGTCGAGCTCGCTCGGGCGCGGGCGGTGCCGC 1258
QY 1201 gagaatgaatgaagg---gggtggggttgatctgagatctgagatagagaggtgtgtggagat 1257
DB 1259 GCCTCCGATGCTGGCCGGCGGCTGTTCCTTCGGGGGAAGTGGAGCGGTGCTGCGCGC 1318
QY 1258 ggtgatggatgggggtgagaagagcaaaactagtagagagaaaaatgcataaaatgggaagac 1317
DB 1319 CGTCATGAGCGGGGGAGCGCGCTCTGCTGCACGCAAGCGCGCGGGGGAATGGAGGGA 1378
QY 1318 ttggtccagaagaagcattgggagagagatggatcttcaactcaagaatactcaagcccttct 1377
DB 1379 CAGGGTTCGCGCGGTGGCAACCCGGTGGCAGCTCTGACCCGCAACCTGGACGAGTTCGT 1438
QY 1378 tca 1380

```





```
Db 793 GATGAACACGCGTGTGAATGGCTTGATTCGAAGAAATCAAGTCCATTTGTTATGTT 852
QY 839 tcgtttggagagcttttgaggtttccaaaggccacaaatgaagagattgggaaaggccta 898
Db 853 TGTGTTGGAGTACACAGAGATTTCACACAGACACAGATCGAAGAACTGCTATGGGGCTA 912
QY 899 ttgacctgcggagggcgcttttatgatgatacgagacagaaagaatgacgacggcgaa 958
Db 913 GAAGCCTCTGGACAAAGATTTCATTTGGGTATCAGAACAGGGGAATGAAGATGGCTCCCA 972
QY 959 gaagaagaagaaggttgagttgcattgggggaattgaaaaaatgggaaaaatagtttcg 1018
Db 973 GAAGGATTCGAGGAAAGAAC-----AAAAAGAAAGGTTTAATCATAAGAGGA 1020
QY 1019 tgggtcgcagttgaggtttcttgccgcaccctgcgttgggattttctgtagcgaattgt 1078
Db 1021 TGGGCACCCCAAGTGTGTGATTTGATCAGAAAGCTATTGGAGCTTTTGTACTCATTTGT 1080
QY 1079 gggtagaactcgctgagagcttgagttgcggggttccgggtggtgcggcgag 1138
Db 1081 GGATGAACCTCGACACTGGAAGGAATATCAGCAGGGGTACCAATGTTGACATGCCAGTA 1140
QY 1139 tggttgatcagcagcaatgcgaagctgattgagatcgctggggggacaggggtgaga 1198
Db 1141 TTTGCGGAACAGTTTTCATGAGAAAGTTGGTGAAGTGAAGTGAAGTGAAGTGGT 1200
QY 1199 gtgagaatgaagaagg 1216
Db 1201 GTTGGTTCGAAGCAATGG 1218
```

## RESULT 5

```
US-08-522-421-1
; Sequence 1, Application US/08522421
; Patent No. 5908973
; GENERAL INFORMATION:
; APPLICANT: Abu-Bakar, Umi Kalsom
; APPLICANT: Barton, Sarah Louise
; APPLICANT: Gallego-Velgas, Pedro Pablo
; APPLICANT: Gray, Julie Elizabeth
; APPLICANT: Grierson, Donald
; APPLICANT: Lowe, Alexandra Louise
; APPLICANT: Picton, Steve
; APPLICANT: Whotton, Lee Collin
; TITLE OF INVENTION: DNA, DNA CONSTRUCTS, CELLS AND PLANTS
; TITLE OF INVENTION: DERIVED THEREFROM
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; FILING DATE: 11-JAN-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305868.3
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305869.1
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305859.2
; FILING DATE: 22-MAR-1993
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305865.9
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305866.7
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305867.5
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305860.0
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305862.6
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9314351.9
; FILING DATE: 12-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9320988.0
; FILING DATE: 12-OCT-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: ERTL1B
; US-08-522-421-1

Query Match 4.4%; Score 66.8; DB 3; Length 1669;
Best Local Similarity 50.0%; Pred. No. 5.5e-09;
Matches 195; Conservative 0; Mismatches 192; Indels 3; Gaps 1;

QY 779 gagagaataaactgcgtgagtgattggacacgaagccgaactcttcggtggtatgtg 838
Db 732 GAAGACAATAGCTGCATCGAGTGGCTCGATAGACAAACCCCTTAACCTCTCTATATGC 791
QY 839 tcgtttggagagcttttgaggtttccaaaggccacaaatgaagagattgggaaaggccta 898
Db 792 AGCTTGGGTAGCCTAGTAGGAGATTGATCACAAGAGTTGATTGAGACTGCTTGGGGATTA 851
QY 899 ttgacctgcggaaaggcgcttttatgatgatacgagacagaaagaatgacgacggcgaa 958
Db 852 GCTAATAGCGCATCAACCGTTCTTGTGGGTTATTTCGACCTGGCTCTGTCTCTGCTTTCAA 911
QY 959 gaagaagaagaaggttgagttgcattgggggaattgaaaa---aaatggggaaaaatagt 1015
Db 912 TGTGCTGAGGCACCTGCCTGATGTTTGTAGAAAATGGTAGGAGAAAGAGACCAATAAGTG 971
QY 1016 tcgttgctcgcagttggaggttctgcccacccctgcgttgggattgttcgtgaagcat 1075
Db 972 AAATGGGCACCAAAAAACAGGTGCTTGACATCCCGCGGTAGCAGGGTTTTTCACTCAT 1031
QY 1076 tgggttgaaactcggctgtggagagcttgagttgggggttcgggtggtggtggtgcg 1135
Db 1032 TGTGGTTGGAATTTCTACGCTTGAAAGTATATGTGAAGAACTCCCTATGTTGGTGTGAGGCCA 1091
QY 1136 cagtgtttgatcagacgacgaatgcgaag 1165
Db 1092 TTTCTAGCAGACCAACTGGTGAACGCNAGG 1121
```

## RESULT 6

```
US-08-797-226-1
; Sequence 1, Application US/08797226
; Patent No. 5959180
; GENERAL INFORMATION:
; APPLICANT: MOEHS, CHARLES P
; APPLICANT: ALLEN, PAUL V
```





```
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-258-261B-6

Query Match          3.1%; Score 46.6; DB 1; Length 28958;
Best Local Similarity 47.4%; Pred. No. 0.0078;
Matches 176; Conservative 0; Mismatches 189; Indels 6; Gaps 1;

Qy 144 ggccgcatggccaaacacagcctccgctgcccgaacccacccggcctcgacttcg 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10208 GCCACAGCCAGGCGGAGATCGCCGCCGCTTCGTGCGAGGCGCTCTCTCCCTCGAGGACG 10267

Qy 204 tggcggttcgcagggctacgacgagggctgaagccctgcccggacggggaagcgctaca 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10268 CGGCGCGATCGCGCCCTCGCGAGGAAAGCGCTCACCACCGTCGCGCGCAACGCGGCA 10327

Qy 264 tgcgagatgaagcccgccgctccgagggccttaagaacctcttcaacaaccacg 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10328 TGGCGCGCTCGAGTCGGCGCTCCGACCTCCAGACCTACCTCGCTCCCTGGGCGGCA 10387

Qy 324 acgtcacgttcgtctactccacccctcttgcctggcgccgaggtggcgcggtgagt 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10388 GGCTCTCCACCGCCCGCTCAACACCCCA-----GGGCTACCTCTGTATCCGCGGAGC 10441

Qy 384 cccaggtcccgagccctctctgtggtcgagcccgccacccgctgtgtgcataattact 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10442 CCGCGCGCTCGAGCGCTGTCTCGACGTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGA 10501

Qy 444 tctactcaacggctacgacgagatcgacgcccgttcgagagaaattcagctccctc 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10502 TCCGGCTCGACTACGCTCCCACTCCGCCAGATGGACGCGCTCCAGACGAGCTCGCG 10561

Qy 504 ggttcacacc 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10562 CAGGTCTAGCC 10572

RESULT 10
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
```

```
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-456-837-6

Query Match          3.1%; Score 46.6; DB 1; Length 28958;
Best Local Similarity 47.4%; Pred. No. 0.0078;
Matches 176; Conservative 0; Mismatches 189; Indels 6; Gaps 1;

Qy 144 ggccgcatggccaaacacagcctccgctgcccgaacccacccggcctcgacttcg 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10208 GCCACAGCCAGGCGGAGATCGCCGCCGCTTCGTGCGAGGCGCTCTCTCCCTCGAGGACG 10267

Qy 204 tggcggttcgcagggctacgacgagggctgaagccctgcccggacggggaagcgctaca 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10268 CGGCGCGATCGCGCCCTCGCGAGGAAAGCGCTCACCACCGTCGCGCGCAACGCGGCA 10327

Qy 264 tgcgagatgaagcccgccgctccgagggccttaagaacctcttcaacaaccacg 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10328 TGGCGCGCTCGAGTCGGCGCTCCGACCTCCAGACCTACCTCGCTCCCTGGGCGGCA 10387

Qy 324 acgtcacgttcgtctactccacccctcttgcctggcgccgaggtggcgcggtgagt 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10388 GGCTCTCCACCGCCCGCTCAACACCCCA-----GGGCTACCTCTGTATCCGCGGAGC 10441

Qy 384 cccaggtcccgagccctctctgtggtcgagcccgccacccgctgtgtgcataattact 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10442 CCGCGCGCTCGAGCGCTGTCTCGACGTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGA 10501

Qy 444 tctactcaacggctacgacgagatcgacgcccgttcgagagaaattcagctccctc 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10502 TCCGGCTCGACTACGCTCCCACTCCGCCAGATGGACGCGCTCCAGACGAGCTCGCG 10561

Qy 504 ggttcacacc 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10562 CAGGTCTAGCC 10572

RESULT 11
US-08-457-342-6
```











**THIS PAGE BLANK (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2000, 21:13:28 ; Search time 3616.02 Seconds  
(without alignments)  
1837.825 Million cell updates/sec

Title: US-09-147-955-1  
Perfect score: 1507  
Sequence: 1 gaaattcccaaaaatgg.....ttataatttagtaacaaaa 1507

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: gb\_est1:\*
- 21: gb\_est2:\*
- 22: gb\_est3:\*
- 23: gb\_est4:\*
- 24: gb\_est5:\*
- 25: gb\_est6:\*
- 26: gb\_est7:\*
- 27: gb\_est8:\*
- 28: gb\_est9:\*
- 29: gb\_est10:\*
- 30: gb\_est11:\*
- 31: gb\_est12:\*
- 32: gb\_est13:\*
- 33: gb\_est14:\*
- 34: gb\_est15:\*
- 35: gb\_est16:\*
- 36: gb\_est17:\*
- 37: gb\_est18:\*
- 38: gb\_est19:\*
- 39: gb\_est20:\*
- 40: gb\_est21:\*
- 41: gb\_est22:\*
- 42: gb\_est23:\*
- 43: gb\_est24:\*

- 44: gb\_est25:\*
- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*
- 54: em\_est22:\*
- 55: em\_est23:\*
- 56: em\_est24:\*
- 57: em\_est25:\*
- 58: em\_est26:\*
- 59: gb\_est33:\*
- 60: gb\_est34:\*
- 61: gb\_est35:\*
- 62: gb\_est36:\*
- 63: gb\_est37:\*
- 64: gb\_est38:\*
- 65: em\_est27:\*
- 66: em\_est28:\*
- 67: em\_est29:\*
- 68: em\_est30:\*
- 69: gb\_est39:\*
- 70: gb\_est40:\*
- 71: gb\_est41:\*
- 72: gb\_est42:\*
- 73: gb\_est43:\*
- 74: gb\_est44:\*
- 75: em\_est31:\*
- 76: em\_est32:\*
- 77: em\_est33:\*
- 78: em\_est34:\*
- 79: gb\_est45:\*
- 80: gb\_est46:\*
- 81: gb\_est47:\*
- 82: em\_est35:\*
- 83: em\_est36:\*
- 84: em\_est37:\*
- 85: gb\_est48:\*
- 86: gb\_est49:\*
- 87: gb\_est50:\*
- 88: gb\_est51:\*
- 89: gb\_est52:\*
- 90: gb\_est53:\*
- 91: gb\_est54:\*
- 92: gb\_est55:\*
- 93: gb\_gss1:\*
- 94: gb\_gss2:\*
- 95: gb\_gss3:\*
- 96: gb\_gss4:\*
- 97: em\_gss1:\*
- 98: em\_gss2:\*
- 99: em\_gss3:\*
- 100: em\_gss4:\*
- 101: gb\_gss5:\*
- 102: gb\_gss6:\*
- 103: gb\_gss7:\*
- 104: gb\_gss8:\*
- 105: gb\_gss9:\*
- 106: em\_gss5:\*
- 107: em\_gss6:\*
- 108: em\_gss7:\*
- 109: em\_gss8:\*
- 110: em\_gss9:\*
- 111: em\_gss10:\*
- 112: em\_gss11:\*
- 113: gb\_gss10:\*
- 114: gb\_gss11:\*
- 115: em\_gss12:\*
- 116: gb\_gss12:\*

117: gb\_gss13:\*  
118: gb\_gss14:\*  
119: gb\_gss15:\*  
120: gb\_gss16:\*  
121: gb\_gss17:\*  
122: gb\_gss18:\*  
123: gb\_gss19:\*  
124: em\_gss13:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	261.6	17.4	647	40	AI488782 EST247121
2	237.2	15.7	536	44	AI771830 EST252930
3	219.2	14.5	546	45	AI897621 EST267064
4	214.2	14.2	747	71	AK349414 GM310007A
5	211.8	14.1	559	45	AI896013 EST265456
6	209.2	13.9	478	45	AI899102 EST268545
7	199.2	13.2	520	40	AI487571 EST245893
8	194.4	12.9	649	72	AW442098 EST311494
9	191.2	12.7	500	40	AI489014 EST247353
10	188.4	12.5	585	63	AW034683 EST278414
11	176.8	11.7	476	40	AI483541 EST249390
12	167.4	11.1	517	44	AI779099 EST259978
13	162.2	10.8	658	79	AW651280 EST329734
14	150.8	10.0	612	63	AW034633 EST278317
15	149.8	9.9	566	45	AI896026 EST265469
16	149.4	9.9	602	63	AW031351 EST274805
17	149.4	9.9	605	69	AW220656 EST297125
18	149.2	9.9	518	63	AW032936 EST276495
19	149.2	9.9	534	72	AW442216 EST311612
20	149.2	9.9	558	63	AW030266 EST273521
21	149.2	9.9	576	63	AW031286 EST274661
22	149.2	9.9	579	63	AW032414 EST275953
23	149.2	9.9	606	63	AW035724 EST281878
24	149.2	9.9	613	63	AW035364 EST280926
25	149.2	9.9	642	79	AW651250 EST329704
26	149.2	9.9	644	69	AW220874 EST297343
27	147.6	9.8	591	79	AW648641 EST327191
28	147.6	9.8	596	69	AW222352 EST299163
29	145.8	9.7	540	44	AI775659 EST256759
30	145.6	9.7	518	63	AW030636 EST273891
31	144.2	9.6	534	63	AW031171 EST274709
32	141.2	9.4	650	63	AW030131 EST273386
33	140.6	9.3	467	45	AI896353 EST265796
34	139.8	9.3	548	44	AI778526 EST259405
35	136.8	9.1	627	71	AW398421 EST298268
36	136.2	9.0	619	72	AW459541 SH42H03.Y
37	134	8.9	513	45	AI895821 EST265264
38	134	8.9	519	44	AI772231 EST253331
39	133.8	8.9	544	69	AW216889 EST295603
40	132.6	8.8	466	62	AW440325 AV440325
41	132	8.8	537	79	AW650189 EST328643
42	131	8.7	534	63	AW030442 EST273697
43	127.2	8.4	435	79	AW650188 EST328642
44	125	8.3	523	47	AI997552 701668943
45	124	8.2	287	64	AW127679 M110425 D

# ALIGNMENTS

RESULT 1  
AI488782  
LOCUS AI488782 647 bp mRNA EST 29-JUN-1999  
DEFINITION EST247121 tomato ovary, TAMU Lycopersicon esculentum cDNA clone  
CLED18D16, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AI488782  
AI488782.1 GI:4384153  
EST  
tomato.  
Lycopersicon esculentum

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; eunasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 647)

Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Llang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.  
Generation of ESTs from tomato carpel tissue  
Unpublished (1999)

TITLE  
JOURNAL  
COMMENT

On Feb 18, 1999 this sequence version replaced gi:4297434.  
Contact: David Frisch  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4365  
Fax: 864 656 4293  
Email: dfrisch@clemson.edu

FEATURES  
source

Location/Qualifiers  
1..647  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone\_lib="tomato ovary, TAMU"  
/tissue\_type="carpel"  
/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5', and 3' ends located at the EcoRI and XhoI sites, respectively."  
224 a 97 c 141 g 185 t

BASE COUNT  
ORIGIN

Query Match 17.4%; Score 261.6; DB 40; Length 647;  
Best Local Similarity 67.3%; Pred. No. 8.9e-63;  
Matches 404; Conservative 0; Mismatches 184; Indels 12; Gaps 2;

QY 579 tgaagagagagctggaacttttagacggtgaagagagagagctgtggtgaacacgt 638

DB 57 TCAAGAGCAACTAGACACATTAGATGGTGAAGAAATCCAAAGGTACTTGTGAATACAT 116

QY 639 ttgatcggttgagcccgccactcacgctattgataggtatgattgatcggtatcg 698

DB 117 TTGATGCATTAGCTAGACCTAGACCCACTCAAGCTATTGGAATAATCAATTTAATGGAATTG 176

QY 699 ggccgttgattccctccgctcttcttgagcgggagatccctccgaaacgtcttacggcg 758

DB 177 GACCAATGATCTCTTCATCATCTCTGGTGGAAAGATTCATTGGATCTCTCATTTGGTG 236

QY 759 ggcattcttcgaaataatcgaggagagataactgcgtgagtggttggacacgagccga 818

DB 237 GTGATCTTTTCAAAAGTCAAAATG---ATGACTACATGGAATGGTTAAACACAAAGCCTA 293

QY 819 aatcttcggtggtgatgtctgttttgagcgttttgaggtttccaaaggcacaatgg 878

DB 294 AATCATCAATTGTTATATCTCATTTGGGAGTCTATTGAATTTATCAAGAAACCAAGG 353

QY 879 aagagattgggaaaggctattgacctcggaagggcgcttttttatgattgatacagagac 938

DB 354 AGGAGATTGCAAAAGGTTTGATAGACATCCAAAGGCCATCTTATGGTAAATAAGAGATC 413

QY 939 aagaagaatgacacgagcgaagagaagaagagagcttgagttgagtgaggaataa 998

DB 414 AA-----GAAGAAGAAAAAAGAGAGAAAATTAAGTTGCATGATGGAATTAGAGA 464

QY	853	tttgagggtttccaaagccacaaatggaagagattggaaagggtctattagcctgcggaag	912		
Db	124	ATTGGAATTTATCAAGAAACCAAAAGGAGGAGATTGCAAAAAGGGTTGATAGAGATCCAAAG	183		
QY	913	qccgtttttatgatgatacagaacaagaagaatgacgacgagcgaagaagaagaagaaga	972		
Db	184	GCCATCTTTATGGGTAATAAG-----ATCAAGAAGAAGAAAAGGAAGAGAA	234		
QY	973	gttgagttgcattggggaattgaaaaaaatgggggaaatagtcttcgtgggtgcgcagtt	1032		
Db	235	ATTAAGTTGTCATGATGAATTAGAAGAACGAAGGAAAAATAGTACCATGGTGTTCACAACT	294		
QY	1033	ggaggttctgggcacacctcgttgggatgttttctgtgacgcattgtgggtggaactcggc	1092		
Db	295	TGAAGTCCTGACACATCCATCTTTAGGATGTTTGTCTCGACACTGGGATGGAATTCGAC	354		
QY	1093	tgtggagagcttgagttgcggggtccggtgggtggcggtgcccgcagttggttgcacagac	1152		
Db	355	TCTAGAAAGCCTATCGTCAGGTGTACTGTGGTGGGCATTTCTCTATTTGGACAGATCAAGG	414		
QY	1153	gacgaatgcgaagctgattgagatgcgtgggggacaggggtgagtgagatggaatgaatga	1212		
Db	415	GACTAATGCAAGTTGATTGAAGATGTTTGGACACAGGTGCTCAGGATGAGAGTGAATGA	474		
QY	1213	agggggtggggttgatggatctgagatagagagagtggtgagagatggtgagtgagtg	1272		
Db	475	AGATGGTGTGTTTGAGAGTGTATGAATATAAAGGTGCATAGAAATTTGATGATGTTG	534		
QY	1273	tg 1274			
Db	535	AG 536			
RESULT	3				
LOCUS	AI897621				
DEFINITION	EST 546 bp mRNA EST 27-JUL-1999				
ACCESSION	EST267064	tomato ovary, TAMU Lycopersicon esculentum cDNA clone			
VERSION	CLED28K21	mRNA sequence.			
KEYWORDS	AI897621				
SOURCE	AI897621.1	GI:5603523			
ORGANISM	tomato.				
REFERENCE	Lycopersicon esculentum				
AUTHORS	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.				
TITLE	1 (bases 1 to 546)				
	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.V., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.				
JOURNAL	Generation of ESTs from tomato carpel tissue				
	Unpublished (1999)				
COMMENT	On Apr 7, 1998 this sequence version replaced gi:3035329.				
	Contact: David Frisch				
FEATURES	Clemson University Genomics Institute				
	Clemson University				
source	100 Jordan Hall, Clemson, SC 29634, USA				
	Tel: 864 656 4366				
Location/Qualifiers	Fax: 864 656 4293				
	Email: dfrisch@CLEMSON.EDU				
prime sequence.	5 prime sequence.				
	1. .546				
prime sequence.	/organism="Lycopersicon esculentum"				
	/cultivar="TA96"				
prime sequence.	/db_xref="taxon:4081"				
	/clone="cLED28K21"				
prime sequence.	/clone_lib="tomato ovary, TAMU"				
	/tissue_type="carpel"				
prime sequence.	/dev_stage="5 days pre-anthesis to 5 days post-anthesis"				
	/lab_host="X11-Blue MRF"				

/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2: XhoI; cLED - Tomato Carpel EST Library, OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."

BASE COUNT 196 a 82 c 114 g 154 t  
ORIGIN

Query Match 14.5%; Score 219.2; DB 45; Length 546;  
Best Local Similarity 67.5%; Pred. No. 6.9e-51;  
Matches 343; Conservative 0; Mismatches 153; Indels 12; Gaps 2;

QY 579 taaagagagctgaacatttagacggtgagagagcggaagtgttggaacact 638  
DB 50 TCAAGAGCACTAGACACATTAGATGGTGAAGAAATCCAAGGTACTTGTGAATACAT 109  
QY 639 ttgatgcgttgagcccgatgcactcacggtctattgataggtatgattgacggatcg 698  
DB 110 TTGATGCATTAGAGCTAGAGCCACTCAAGCTATTGAAAAATACAATTTAATTGGAATTG 169  
QY 699 ggcgttgattccctcgctctcttgacgagcggtgagatccctccgaaacgtcttacggcg 758  
DB 170 GACCATTGATCTCTCATCATCTTCTGGGTGGAAGAGATTTCATTGGAATCTTCAATTGGTG 229  
QY 759 gcgattcttcgaaaaaatcgaggagaaataactcggtgagtggttggtggacacgaagcga 818  
DB 230 GTGATCTTTTCAAGTCAATG---ATGACTACATGGAATGGTTAAACACAAAGCCTA 286  
QY 819 aatcttcggtgtatgtgtgctgttggtggagcgcttttgaggtttccaaaggcacaatgg 878  
DB 287 AATCATCAATTGCTTATATCTCATTTGGGAGTCTATTGAATTTATCAAGAACCAAGG 346  
QY 879 aagagattgggaagggctattagctcgggaagcgctttttatgtagatgacagaaac 938  
DB 347 AGGAGATTGCAAAAGGGTGTATAGAGATCCAAAGGCCATTCTTATGGGTAAAGAG--- 403  
QY 939 agaagaatgacacgagcggaagaagaagagagtggtgagtcattggtgggaattgaasa 998  
DB 404 -----ATCACCAGAGAGAAAAGAGAGAGAAATTCAGTTGCATGATGGAATTAGAGA 457  
QY 999 aaatggggaaaaatagttcgtggtgtcgcagttggaggttctggtgcacccctggttgg 1058  
DB 458 AGCAAGGGAATAATAGTACCATTGGTTCACAACTTGAAGTCTGACACATCATCTTTAG 517  
QY 1059 gatgttcgtgacgaattgtggtggaa 1086  
DB 518 GATGCTTCTGCTCGCACTGTGGATGGA 545

RESULT 4  
AW349414/c  
LOCUS  
DEFINITION GM210007A20D2R Gm-r1021 Glycine max cDNA 3', mRNA sequence.  
ACCESSION AW349414  
VERSION AW349414.1 GI:6847124  
KEYWORDS EST.  
SOURCE soybean.  
ORGANISM Glycine max  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.  
REFERENCE 1 (bases 1 to 747)  
AUTHORS Vodkin,L., Kelm,P., Shoemaker,R., Retzel,E., Khanna,A., Corvelli,V., Erpelidng,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.  
TITLE A Functional Genomics Program for Soybean (NSF 9872565)  
JOURNAL Unpublished (1999)  
COMMENT On Jul 9, 1999 this sequence version replaced q1:5434916.  
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565).  
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics  
University of Illinois

Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA  
Tel: (217) 244-6147  
Fax: (217) 333-4582  
Email: l-vodkin@uiuc.edu  
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888)919-3324 or (314) 427-3324 or contact:clones@genomesystems.com or info@genomeystems.com web site:www.genomesystems.com  
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES  
Location/Qualifiers  
1..747  
source  
/organism="Glycine max"  
/cultivar="Williams"  
/db\_xref="taxon:3847"  
/clone\_lib="Gm-r1021"  
/tissue\_type="root"  
/lab\_host="XL10-Gold"  
/note="Vector: pBluescript II XR; Site.1: EcoRI; Site.2: XhoI; Library Gm-r1021 is a sequence-driven, rereacked set of the original library Gm-cl004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. StrataGene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. The Gm-cl004 library was constructed by Dr. Paul Kelm & Virginia H. Corvelli, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.kelemenau.edu, virginia.corvelli@uau.edu. The contig analysis to select unique genes was performed by the Laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html . Rereacking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html."

BASE COUNT 206 a 235 c 97 g 187 t 22 others  
ORIGIN

Query Match 14.2%; Score 214.2; DB 71; Length 747;  
Best Local Similarity 59.7%; Pred. No. 1.9e-49;  
Matches 414; Conservative 0; Mismatches 243; Indels 36; Gaps 3;

QY 707 attcctccgctcttgacgagcgagatccctccgaaacgtcttacgagcgagatctt 766  
DB 733 ATTCNNNNNCNTTCNNNGNCGTAAAGATCCCGCGGATACTNNNTATGCGGTGATNNN 674  
QY 767 ttccgaaaaatcgaggagaataactcggtgagtggttgacacgaagcgaatcttcg 826  
DB 673 NNCNNTGCTTC-----AAATGATTACGTTGTAAGTTGGACTCACAGCCTGAGTTATCT 620  
QY 827 gtgggtatgtgtcgtttgggagcggtttgaggtttccaaaggcacaatggaagagatt 886  
DB 619 GTGGTTTATGTTTCATTGTTGTTACCTTGTGCTGTTGGCTGATAGACAGATGAAGAGCTT 560  
QY 887 gggaaagggctattagcctcggaagggcggtttttatgatgatgacgaaacagaagat 946  
DB 559 GCACGCGCGCTGTAGATTCCGATATCTCTTCTTGTGGTTCATTAGAGATATG----- 504  
QY 947 gacgacggcgaagaagaagaagaagattgattgcattgggaattggaaatgggg 1006  
DB 505 -----CAAGGAATAGAAATAGATACTGCAGAGAGAGAACTGGAGCAGAGGGGT 461  
QY 1007 aaaaatagttctggtgtcgtccgagttggaggttcttcgagcaccctcgctgggagatttc 1066  
DB 460 AAGATTGTGAATGGTGTCTCAGGTGGAGGTCTGTCGCATGGTTCTTGGTGGTCTTTT 401  
QY 1067 gtgacgcatgtgggtggaactcggtctgagagagctttagttgaggggttccggtggtg 1126  
DB 400 GTAACGCATTGTGGTGAATTCGACTATGGAAGTTTGGGTTCGGGGGTTCCTATGTTG 341



```

/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage=5 days pre-anthesis to 5 days post-anthesis"
/lab_host="X11-Blue MRP"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED - Tomato Carpel EST library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5',
and 3' ends located at the EcoRI and XhoI sites,
respectively."
172 a 67 c 102 g 137 t

```

	Query Match	13.9%	Score 209.2	DB 45	Length 478
	Best Local Similarity	67.3%	Pred. No. 4.3e-48		
	Matches 330	Conservative 0	Mismatches 148	Indels 12	Gaps 2
QY	606	gtgaagaagagggaagattgttgtaaacggtttgatcgcttgagccgagccgcatcactca	665		
Db	1	GTGAAGAAAATCCAAAGGTACTTGTGAATCATTTGTATGATAGCTAGAGCCACTCA	60		
QY	666	cggctattgataggtatgattcggtgacggcgcttgattccctccgcctctcttgg	725		
Db	61	AAGCTATTGAAAATAACAATTAATTGGAATTCGACCATTCATTCCCTCATCATCTCTGG	120		
QY	726	acggcgagatccctccgaacgctcttcacggcgcgatcttttcgaaaaatcgagagaga	785		
Db	121	GTGAAAAGATTCATTGGAAATCTTCATTGTGTGTGATCTTTTTCAAAAGTCAAAATG	177		
QY	786	ataactcgttggaggtttggacacgaagccgaataatttcgggtggtgtatgtgcgtttg	845		
Db	178	ATGACTACATGGAAATGTTAAACAAGGCTAAATCATCAATGTTTATATCTCATTTG	237		
QY	846	ggagcgcttttgaggtttcccaaggccacaaattggagagattgggaaggcgctatagcct	905		
Db	238	GGAGCTATTGAAATTTATCAAGAAACCAAGAGGAGAGATTGCAAAAGGTTGATAGAGA	297		
QY	906	gcggaaagccgctttttatgtagatgacgagaacagaagaatgacgacggcgaaagaag	965		
Db	298	TCCAAAGGCCATCTTATGGTATAAGAG-----ATCAAGAAGACAAAAGAAAG	348		
QY	966	aagaagagttgagttgcatggggaattgaaaaaaatgggaaaaatagtttcgggtgct	1025		
Db	349	AAGAGAAATTAAGTTGCATGATGGAAATTAGAGAAAGGGAATAATAGTACCATGGTGTT	408		
QY	1026	cgcagttggaggttcttcgacaccctcgcttgggatttcttcgtgacgcatctgtgggtgga	1085		
Db	409	CACAACTTGAAGTCTGCACATCCATCTTTAAGATGTTTGTCTCGCACTGTGGATGGA	468		
QY	1086	actcggctgt	1095		
Db	469	ATTCGACTCT	478		

RESULT	7				
LOCUS	AI487571	520 bp	mrna	EST	29-JUN-1999
DEFINITION	EST245893 tomato ovary, TAMU Lycopersicon esculentum cDNA clone				
	CDED10D21, mRNA sequence.				
ACCESSION	AI487571				
VERSION	AI487571.1	GI:4382942			
KEYWORDS	EST.				
SOURCE	tomato.				
ORGANISM	Lycopersicon esculentum				
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; easterids I; Solanales; Solanaceae; Solanum; Lycopersicon.				
REFERENCE	1 (bases 1 to 520)				
AUTHORS	Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E., Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C., Martin, G.B., Tankale, S.D. and Giovannoni, J.				
TITLE	Generation of ESTs from tomato carpel tissue				

Unpublished (1999)  
On Apr 7, 1998 this sequence version replaced gi:3035667.  
Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU.  
Location/Qualifiers  
1 520

```

FEATURES
  source
    1. .520
      /organism="Lycopersicon esculentum"
      /cultivar="TA496"
      /db_xref="taxon:4081"
      /clone="cLED10D21"
      /clone_lib="tomato ovary, TAMU"
      /tissue_type="carpel"
      /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
      /lab_host="XLI-Blue MRP"
      /note="Vector: pBlueScript SK(-); Site.1: EcoRI; Site.2:
XhoI; cLED - Tomato Carpel EST Library_ OligoDT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT      194 a      76 c      107 g      143 t
ORIGIN

Query Match      13.2%; Score 199.2; DB 40; Length 520;
Best Local Similarity 67.2%; Pred. No. 2.9e-45;
Matches 317; Conservative 0; Mismatches 143; Indels 12; Gaps 2;

```

Qy	579	tgaaggagaagctggaaacatttaaacggttgaagagaaggcgaaaagtgttggtggaacacgt	638
Db	57	TCAAAGAGCAACTAGACACATATTAGATGGTGAAGAAAATCCAAGAGTACTTTGTTGAATCAT	116
Qy	639	ttgatgcgttggagcccgatgcactcagcgctattgatagtagtgaagtgtgatcgggatacg	698
Db	117	TTGATGCATTAGACTAGAGCCACTCAAGCTATTGAAATAACAAATTAAATTCGAATTG	176
Qy	699	ggcgttgtattcccttcgcgcctcttggagcggcgagatccctccgaaacgtctctcaagcg	758
Db	177	GACCATTGATTCCCTTCATCATTCTTGGTGGTGAAGAAGATTCAATCGAATCTTTCATTTGGTG	236
Qy	759	gcgcatcttttcgaaaaatcggagagagaataactcgtggagtggttggaacacgaagccga	818
Db	237	GTGATCTTTTTCAANAAGTCAAAATG---ATGACTACATGGAATGCTTAAACAACAAGCCCTA	293
Qy	819	aatcttcgggtgtatgtcgttctggagcgtttttgagagttccc aaagcacaaatgg	878
Db	294	AATCATCAATTGTTTATATCTCATTTGGAGCTATTGAAATTTCAAGAACAACCAAAAGG	353
Qy	879	aagagattgggaaggcgctattagcctcgcgaaagccgcttttatgatgatatacgagaac	938
Db	354	AGGAGATTGCCAAAAGGGTTGATAGAGATCCAAGAGCCATCTTATTGGGTATAAAGAG---	410
Qy	939	agaagaatgatcgcgcgcgaagaagaagaagsgttgagttgcatcggggaaattgaaa	998
Db	411	-----ATCANGAGAGAGAAAAAGAGNAGAGAAATTAAGTTGGCATGATGGAATTAGAGA	464
Qy	999	aatcggggaaaaatagtcttcgtggtcgtcgcagttggagggtcttcggcgaccoc	1050
Db	465	AGCAAGGGAAAAATAGTACCATGGTGTTCACAACCTTGAACTCCTCACACATCC	516

RESULT	8
AW442098	
LOCUS	AW442098 649 bp mRNA EST
DEFINITION	EST311494 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA clone CLEN21C12 5', mRNA sequence.
ACCESSION	AW442098
VERSION	AW442098.1 GI:6977349
KEYWORDS	EST.







Martin, G.B., Tankley, S.D. and Giovannoni, J.J.  
Generation of ESTs from tomato carpel tissue  
Unpublished (1999)  
Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU.

**FEATURES**  
**source**

```

1. /organelle="Lycopodium esculentum"
2. /cultivar="TA436"
3. /db_xref="taxon:4081"
4. /clone="cLE025M13"
5. /clone_lib="tomato ovary, TAMU"
6. /tissue_type="carpel"
7. /dev_stage="5 days pre-anthesis to
8. /lab_host="X11-Blue WRF"
9. /note="Vector: pBlueScript SK(-); S.
10. Xhol; cLE0 - Tomato carpel EST Library
11. directionally cloned cDNA in vector
12. and 3' ends located at the EcoRI and
13. respectively."

```

	180 a	64 c	101 g	131 t
BASE COUNT				
ORIGIN				

Query Match 11.7%; Score 176.8; DB 40; Length 476;  
Best Local Similarity 66.7%; Pred. No. 5.5e-39;  
Matches 288; Conservative 0; Mismatches 132; Indels 12;

	Matches	288; Conservative	0; Mismatches	132; Indels	12; Gaps	2;
Qy	579	tgaaggaaagctggaaacttttagacggtgaagagaaggcgaagctgttggtgaacacgt	638			
Db	57	TCAAAGAGCAACTAGACACATTAGATGCTGAAGAAAAATCCAAGGTA	116			
Qy	639	ttgaatcggttggagcccgatgcactcgcgctattgataggatgatgattcgcggatcg	698			
Db	117	TTGATGCAATTAGAGCTAGGCCATCAAGGCTATTGAAAAATACAAATTAATGGAATTG	176			
Qy	699	ggcggttgattccctccgcctctttggacgcgagatccctccgaaacgtctctacggcg	758			
Db	177	GACCAATTGATTCCTTCATCATTCTTGGTGGAAAGATTCATGGAATCTTCATTGGTG	236			
Qy	759	gcgactcttttcgaaaaatcggagagaaataactcgcgtggagtggttgacacgaagccga	818			
Db	237	GTGATCTTTTTTCAAAGCTCAAATG--ATGACTACATGGAATGSGTTAAACACAAGCCGTA	293			
Qy	819	aattctcgggtggtatgctgcgtttggagcgcttttgaggtttccaaagcgcaaatcgg	878			
Db	294	AATCATCAATTTGTTTATCTCATTTGGAGTCTATTGAATTTATCAAGAAACCAAAAGG	353			
Qy	879	aagagattggaaaaagggtctattagcctcgcggaaaggccgctttttagatgatcagaaac	938			
Db	354	AGGAGATTGCAAAAGGGTTGATAGAGATCCAAGAGCCATCTCTATGGGTAAATTAGAG---	410			
Qy	939	agaagaatgacgcgcggcgaagaaagaaagaggttgatgctggtggaaattgaaaa	998			
Db	411	-----ATCAAGAGAGAGAAAAAGAGAGAGAAATTAAGTTGGATGATGGAATTAGACA	464			
Qy	999	aaatggggaaaa	1010			
Db	465	AGCCAGGGCAAAA	476			

**RESULT 12**

AI779099	AI779099	517 bp	mRNA	EST	29-JUN-1999
LOCUS	EST59978	tomato susceptible,	Cornell Lycopersicon esculentum	cDNA	
DEFINITION	clone cLEST19,	mRNA sequence.			
ACCESSION	AI779099				
VERSION	AI779099.1	GI:5277140			

KEYWORDS  
SOURCE  
ORGANISM

## REFERENCE AUTHORS

TITLE	JOURNAL	COMMENT

**FEATURES**

[illegible]

Query Match 11.1%; Score 167.4; DB 44; Length 517;  
Best Local Similarity 68.4%; Pred. No. 2.5e-36;  
Matches 247; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

Qy	579	tgaaggagaagctggaaacttttagacggtgaagaagcgcaagtgttggtgtaaacacgt	638
Db	152	TCAAAGAGCAACTAGACACATTAGATGTTGAGAAAAATCCAAAGGTACTTGTGAATFACAT	211
Qy	639	ttgatgcgttggagcccgatgcactcacgcgtattgatagggtatgagttgatcggatcg	698
Db	212	TTGATGCATTAGAGCTAGAGGCCACTCAAAGCTATTGAAAATAACAATTAATTAATGGAAATTG	271
Qy	699	ggcgttgattccccttcgccttcttggacgcgagatccctccgaaaagctcttacggcg	758
Db	272	GACCATTGATTCCTTCATCATCTTTGGGTGGAAAGAATCATTGGAACTCTTCATTTGGTG	331
Qy	759	gcgactcttcgaaaaactcgaggagaaataactgcgtggsgtgttggacacgaagccga	818
Db	332	GTGATCTTTTTCAAAGTCAAAATG--ATGACTACATGGGAATGTTTAACACAAAAGCCCTA	388
Qy	819	aatcttcgggtgtatgtctgtcttgggagcgcttttgaggtttccaaggcacaatatgg	878
Db	389	AATCATCAATGTTTATATCTCATTTGGGAGTCTATTGAATTTATCAAGAAACCAAAAGG	448
Qy	879	aagagatgggaaaggcgtattagcctgcggaagcgcogtttttcatggtatgcagaaac	938
Db	449	AGGAGATTGCAAAAGGGTTCATAGAGATCCAAAGGCCATCTCTTATGGGTAAATAGAGATC	508

QY 939 a 939  
Db 509 A 509

Db	308	GATCTTCCATCTTTTTTTAGTTTCATCTAGCTCAAAAGATGATGAGTATAGTTTGTCTCT	367
QY	575	---atgatgaaggagcctggaaacttttagacggtgaagagagcgaagcgaagtggttgg	630
Db	368	ACCAACATTCANAGAGCACTAGACACATTAGATGGTGAAGAAATCCAAAGGTACTTGT	427
QY	631	gaacacgtttgatcgcttggtggcccgatgcactcaocggtctattgatagggtatgattgat	690
Db	428	GAATACATTTTGATGCATTAGAGCTAGAGCCACTCAAAGCTATTGAAAAATACAATTAAT	487
QY	691	cgggatcgggcccgttgattccctccgcctctcttttgagacggcggaagatccctccgaaagtc	750
Db	488	TGGAATTTGGACCATTTGATTCCTTCATCATTCCTGGGTGAAAAAGATTCATTTGGAATCTTC	547
QY	751	ttacgcgcgcgatcttttcgaaataatcgagagagaaataaactgcgtgagtggtttggacac	810
Db	548	ATTGTTGGTGGTATCTTTTCAAAAGTCNAATG---ATGACTACATGGNATGGTTAAACAC	604
QY	811	gaagccgaatacttcggtggtgtatgtgtcgtttgggagcgcttttgaggtt	861
Db	605	AAAGCCTAAATCATCAATTTGTTATATCTCATTCATTAGGGAGTCTATTGAATTT	655
RESULT	14		
AW034633			
LOCUS	AW034633	612 bp	15-SEP-1999
DEFINITION	EST278317 tomato callus, TAMU Lycopersicon esculentum cDNA clone		
	CLEC33F2, mRNA sequence.		
ACCESSION	AW034633		
VERSION	AW034633.1	GI:5893389	
KEYWORDS	EST.		
SOURCE	tomato.		
ORGANISM	Lycopersicon esculentum		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.		
AUTHORS	1 (bases 1 to 612) Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E., Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.		
TITLE	Generation of ESTs from tomato callus tissue		
JOURNAL	Unpublished (1999)		
COMMENT	On Jul 7, 1999 this sequence version replaced gi:5405910. Contact: David Frisch Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfrisch@CLEMSON.EDU 5 prime sequence.		
FEATURES	Location/Qualifiers		
source	1..612		
	/organism="Lycopersicon esculentum"		
	/cultivar="TA496"		
	/db_xref="taxon:4081"		
	/clone="cLEC33F2"		
	/clone_lib="tomato callus, TAMU"		
	/tissue_type="callus"		
	/dev_stage="25-40 days old"		
	/lab_host="XL1-Blue MRF"		
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses, Tomato Callus EST Library"		
BASE COUNT	166 a	136 c	133 g 177 t
ORIGIN			
Query Match	10.0%	Score 150.8:	DB 63: Length 612:



**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	1363.5	55.1	460	10	Q9Z827		Q9Zr27 perilla fru
2	1320	53.3	461	10	Q9Z825		Q9Zr25 verberna h
3	1307.5	52.8	443	10	Q9Z826		Q9Zr26 perilla fru
4	1207	48.8	467	10	Q93709		Q93709 nicotiana t
5	1133.5	45.8	474	10	O04930		O04930 arabidopsi
6	1090.5	44.1	456	10	Q23270		Q23270 arabidopsi
7	1084.5	43.8	519	10	Q9ZVY2		Q9ZvY2 arabidopsi
8	1060	42.8	455	10	Q9ZVY5		Q9ZvY5 arabidopsi
9	1049.5	42.4	458	10	Q23406		Q23406 arabidopsi
10	809.5	32.7	479	10	Q23401		Q23401 arabidopsi
11	777	31.4	449	10	Q23822		Q23822 arabidopsi
12	766	30.9	475	10	O23402		O23402 arabidopsi
13	762	30.8	460	10	O48676		O48676 arabidopsi
14	761	30.7	484	10	O23400		O23400 arabidopsi
15	733	29.6	449	10	O23820		O23820 arabidopsi
16	727	29.4	456	10	O22182		O22182 arabidopsi
17	712	28.8	438	10	O22183		O22183 arabidopsi
18	645	26.1	453	10	O22186		O22186 arabidopsi
19	592	23.9	481	10	O9ZWJ3		O9ZwJ3 arabidopsi

```

117 AEVARESQVPSALLWVEPATVLCIYFYENG Y----ADETDAGSDIEQLPRLPPEORSL 172
179 PSFLLPYGAGSLRVALPFPKELIDTLDIAETPKIILVNTFDELEPALNAIEGYKYGIG 238
173 PTFLLPETPE-RFRLLAH---KEKLETDGEERAKAKVLVNTFDALPDALTAIDRYELIG 228
239 PLIPSAFLAGNDPLDASFGGDLFONS--NDYMEWLNKSPNSVVYISFGSLMNPISOME 296
229 PLIPSAFLAGNDPSETSYSGDLEFEKSEENNCVETLTKPKSSVVYVSGSVLRFPAQME 288
297 EISGLIDIGRPPFLWIKI--NEKKEENKKGICIEELEKIGKIVPWCQSLEVLKHPSL 354
289 EIGKGLLACGRPPLWIREQKNDGDEE--EELSCIGELKMGKIVSWCSQLEVLALPAL 347
355 GCFVSHCGWNSALESACGVPVAVPQWTDQMTNAKQVEDVWKSQVYVRINEDGVVESEE 414
348 GCFVTHCGWNSAVESLSCGVPVAVPQWFDQTTNAKLIEDAWGTGVRVRNREGGVDGSE 407
415 IKRCIELVMDGGEGBELKRNKAKWELAREAVKEGGSSHKNLKAFIDVAK 466
408 IERCVEWMDGGEKSLVRENALIKWTLAREAMGEGSSLKLNLAFLHVAR 459

```

```

RESULT 2
Q92R25 PRELIMINARY; PRT; 461 AA.
ID Q92R25
AC Q92R25
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE.
GN HGT8.
OS Verbena x hybrida
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Verbenaceae;
OC Verbena.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PETAL;
RX MEDLINE; 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL; AB013598; BAA36423.1; -.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 461 AA; 51347 MW; 6DCBC3B6 CRC32;

```

```

Query Match 53.3%; Score 1320; DB 10; Length 461;
Best Local Similarity 55.7%; Pred. No. 2.2e-88;
Matches 264; Conservative 76; Mismatches 108; Indels 26; Gaps 10;

QY 1 MVOPHVLITFPAGQHINPALQPAKNLVRMGIEVTTSTSIYAQRMDKSIILNAPKGL-N 59
DB 1 MSRAVLLATFPAGQHINPALQPAKLNADIQVTTFTSVYAWRRMSRTAA--GSNGLIN 58
QY 60 FIFPSDGFDEGDSKDPVYMSQLKCGSETVKKIL--TCSENGOPTICLLYSIFLPW 117
DB 59 FVSFSDGDDGLOPDDGKNYMSKRGIKALSDTLAANNVDQKSKITFVYVSHLPAW 118
QY 118 AAEVAREVHIPSSALLWSQATIIDYIFNPHGYEKAMANESNDPNWSIOLP-GLPLETR 176
DB 119 AAKVAREEHLRSALLWTEPATVLDIEFFYENG Y----SDEIDAGSDAIHLPGSLPVAQR 174
QY 177 DLPSPFLP--YGAGKSLRVALPFPKELIDTLDIAETPKIILVNTFDELEPALNAIEGYKF 234
DB 175 DLPSPFLPSTHERFSL-----MKEKLETLGEERPKVLVNSFDALPDALKAIDYEM 228

```

```

QY 235 YGIGPLIPSAFLAGNDPLDASFGGDLFO---NSNDYMEWLNKSPNSVVYISFGSLMNP 291
DB 229 IAIGPLIPSAFLAGNDKOPSDRSGGDLFKEGSDDDCLEWLNTPRSPSVVYVSGSVNT 288
QY 292 ISOMEIEISGLIDIGRPPFLWIKENKKEENKKGICIEELEKIGKIVPWCQSLEVLKH 351
DB 289 KSQMEIARGLLDGCRPFLWVVRV--EGBEVLISCMEELKRVKIVSWCSQLEVLTH 344
QY 352 PSLGCFVSHCGWNSALESACGVPVAVPQWTDQMTNAKQVEDVWKSQVYVRINEDG-VV 410
DB 345 PSLGCFVTHCGWNSTLESISFGVPVAVPQWFDQTNAKLMEDVWRTGVRVRANEESGV 404
QY 411 ESSEIKRCIELVMDGGEGBELKRNKAKWELAREAVKEGGSSHKNLKAFIDV 464
DB 405 DGDEIRARCIIEVMDGGEKSKRLRESAGKWKDLARKAMEEDGSSVNNLKVPFLDEV 458

RESULT 3
Q92R26 PRELIMINARY; PRT; 443 AA.
ID Q92R26
AC Q92R26
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE HOMOLOGUE.
GN PF3R6.
OS Perilla frutescens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Lamiaceae; Perilla.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SHIKUN; TISSUE=LEAF;
RX MEDLINE; 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL; AB013597; BAA36422.1; -.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 443 AA; 49110 MW; D87B5486 CRC32;

```

```

Query Match 52.8%; Score 1307.5; DB 10; Length 443;
Best Local Similarity 56.6%; Pred. No. 1.7e-87;
Matches 260; Conservative 71; Mismatches 99; Indels 29; Gaps 8;

QY 1 MVOPHVLITFPAGQHINPALQPAKNLVRMGIEVTTSTSIYAQRMDKSIILNAPKGL 58
DB 1 MVRVRLATFPAGQHINPALQPAKLLKAGTDVTTFTSVYAWRRMANTASAAAGNPPGL 60
QY 59 NFIPDGFDEGDSKDPVYMSQLKCGSETVKKILITCSENGOPTICLLYSIFLPWA 118
DB 61 DFVAFSDGDDGDLKPGDGKRYMSENKARGSEALRNLL--NDDVTFVYVSHLPAWA 116
QY 119 AAEVAREVHIPSSALLWSQATIIDYIFNPHGYEKAMANESNDPNWSIOLPGLPLETR 178
DB 117 AEVARLSHVPTALLWVEPATVLCIYHFYENG YADEIDAGSNE----IQLPRLPSEORSL 172
QY 179 PSFLLPYGAGSLRVALPFPKELIDTLDIAETPKIILVNTFDELEPALNAIEGYK 233
DB 173 PTFLLP-----ATPERFLMKKEKLETLDEGEERAKVLVNTFDALPDALTAIDRYE 223
QY 234 FYGIGPLIPSAFLAGNDPLDASFGGDLFONS--NDYMEWLNKSPNSVVYISFGSLMNP 291
DB 224 LIIGPLIPSAFLAGNDGDPSETSYSGDLEFEKSEENNCVWLNKSPSVVYVSGSVLRF 283
QY 292 ISOMEIEISGLIDIGRPPFLWIKI--NEKG-KEEENKKGICIEELEKIGKIVPWCQSLE 348

```



```

Db 284 KAQMEETGKGLACRPFLLWIREQKNDGEEBEEELSCIGELKMKGVSWCSOLEV 343
QY 349 LKHPSLGCFVSHGWSALESAGVVPVAFPOWTDQMTNAKOVEDVWKSQVRVINEOG 408
Db 344 LAHPALGCFVTHGWSAVESLSGPIPVAVPOWFDOTTNAKLIEDAWGTGVRVMEGG 403
QY 409 VVESEIKRCIELVMDGEGEELRNKNAKKELAREAV 447
Db 404 GVDGCEIERCEVMVMDGGDKTKLVRENAIKWKTARQAM 442
RESULT 4
P93709 PRELIMINARY; PRT; 467 AA.
AC P93709;
DT 01-MAY-1997 (T-EMBLrel. 03, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE-INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE
DE (EC 2.4.1.121) (INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE).
GN JIGT.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC Nicotiana.
CC [1]
RN SEQUENCE FROM N.A.
RC STRAIN-BY-2; TISSUE-SUSPENSION CULTURE;
RA KOJIMA H., HASHIZUME K., IMANISHI S., NAKAMURA K.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB000623; BAA19155.1; -.
DR MENDEL; 9421; Nicta; 1525; 9421.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 467 AA; 52989 MW; 3726EB31 CRC32;

Query Match 48.8%; Score 1207; DB 10; Length 467;
Best Local Similarity 50.8%; Pred. No. 3.7e-80;
Matches 234; Conservative 87; Mismatches 132; Indels 14; Gaps

QY 5 HVLTITPQAQGHINPALQFAKNLVKMGIEVFTSTSIYAQSRMDEKSLNAPK--GLNFIP 62
Db 9 HVLLALPGGHHNPSLQFSKLLNLGVKVTLSLSLAFNR-----IKNLPKLEGLTFAP 63
QY 63 FSDGDFGDFHSDKPV--FYMSQLKCGSETVKKIILTCSENGOPITCLLYSIFLPWAAEY 121
Db 64 FSDGDFGDFHSDKPV--FYMSQLKCGSETVKKIILTCSENGOPITCLLYSIFLPWAAEY 123
QY 122 AREVHIPSALLWSOPATILDIYFNPHGYEKAMAN--ESNDPNWSIQLPGLLETRDLP 180
Db 124 AKKLHPSLTFWOPATVFDIYFRNFNFYFNKYSQDQ--IIEPLGPSLSSDFFS 181
QY 181 FLPLYGAKGSLVALPFPFKELIDTLDATETPKILVNTFDELEPEALNAIRGYFYGTGPL 240
Db 182 FV--FDDVKSNDAVESIKKQIETLSEENPRILVNTFDEALNALRVANVMVGLPL 239
QY 241 IPSAFLGNDPLDASFGDDLFOQNSNDYMWMLNKPNSVYVYISFGSLMNPISISOMEISK 300
Db 240 IPSFLDEKDKONFFAADMIESENNMEWLDARANKSVIYIAPGSYAEISSQWMEISQ 299
QY 301 GLIDIGRPFLLWIKENEGKEENKKGCIETEELEKIGKIVPWCQSOLBVLKHPSLGCFVSH 360
Db 300 GLKCGRPFLLWIRETLNGEKPE--EKLTCDELEKIGRIYRWCSQMEVLKHSYGCFLTH 358
QY 361 CGWSALESACGVPVAVFPOWTDQMTNAKOVEDVWKSQVRVINEOGVVESEIEKRCIE 420
Db 359 CGWNTLESASGVPVIVACPIWDOICNAKLIOQWIGVIRVANKSGGIKRDDEFQKIE 418
QY 421 LVMDGEGEELRNKNAKKELAREAVKEGSSHKNLKAFIDDVAKG 467
Db 419 IVMGDAEEGELRNKNAKKELAREAVKEGSSHKNLKAFIDDVAKG 465

```

```

RESULT 5
O04930 PRELIMINARY; PRT; 474 AA.
AC O04930;
DT 01-JUL-1997 (T-EMBLrel. 04, Created)
DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE-INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE
DE (EC 2.4.1.121) (INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE).
GN TAGL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
CC [1]
RN SEQUENCE FROM N.A.
RA GILAD A., SCOLNIK P.A., BAR-ZVI D.;
RL Plant Physiol. 113:1004-1004(1997).
RN [2]
RN SEQUENCE FROM N.A.
RA THORNBURG R.W., GRAHAM R.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: UDP-GLUCOSE + INDOLE-3-ACETATE -> UDP +
CC INDOLE-3-ACETYL-BETA-1-D-GLUCOSE.
DR EMBL; U81293; AAB58497.1; -.
DR MENDEL; 16583; Arab; 2543; 16583.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 474 AA; 53740 MW; 35521D70 CRC32;

Query Match 45.8%; Score 1133.5; DB 10; Length 474;
Best Local Similarity 49.0%; Pred. No. 8.6e-75;
Matches 238; Conservative 77; Mismatches 120; Indels 51; Gaps 14;

QY 4 PHVLTITPQAQGHINPALQFAKNLVK--MGIEVFTSTSIYAQSRMDEKSLNAPKGLNFI 61
Db 12 PHFLTITPQAQGHINPSLEAKLAKLAGTISGARVTFASISAYNRR-MFSTENVETLIFA 70
QY 62 FSDGDFGDFHSDKPV--FYMSQLKCGSETVKKIILTCSENGOPITCLLYSI 113
Db 71 TSDGHDGDFGKSSAYSOKSRODATGNFEMRRRGKETLTLEIDNRKQNRPTCCVITI 130
QY 114 FLPWAAEVAAREVHIPSALLWSOPATILDIYFNPHGYEKAMANESNDPNWSIQLPGLPL 173
Db 131 LLTWVAELAREEHLPSALLWVQPVTFVIFVHYFNGVEDALSEMANTPSSSIKLPPL 190
QY 174 ETRDLPSELPLYGAKGSLVALPFPFKELIDTLDATETPKILVNTFDELEPEALNAI--EGY 232
Db 191 TVRDIPSFVSSNNYVAF--LPAFREQIDSLKEEINPKILINTFOELEPEANSVPDNF 247
QY 233 KFYGIGLIPSAFLGNDPLDASFGDDLFOQNSNDYMWMLNKPNSVYVYISFGSLMNP 292
Db 248 KIVPVGPLL-----TLRTD-FSSRGYIEMLDTKADSSVLYVSGFLAVLSK 293
QY 293 SQMEETISGLIDIGRPFLLWV-----KENEGKEENKKGCIETEELEK----IGKIVP 341
Db 294 KQVLELCKALLQSRPFLWITDKSYRNKDEQEKEED-----CISSEKSFDEIGMVVS 348
QY 342 WCSOLEVLPKPSLGCFSVSHGWSALESAGVVPVAFPOWTDQMTNAKOVEDVWKSQVR 401
Db 349 WCDQFRVNLHRSIGCFVTHGWSNTLESLSVSGVPVAVFPOWTDQMTNAKLLEDCKWTKGR 408
QY 402 V--RINEDG--VVESEIEKRCIELVMDGEGEELRNKNAKKELAREAVKEGSSHKNL 457
Db 409 VMEKKEEGVYVVDSEIRRCIEEVM--DKAEEFRGNATRWKDLAAEAVREGGSSFNHL 466
QY 458 KAFIDD 463
|||:|:

```

```
Db 467 KAFVDE 472
RESULT 6
O23270 PRELIMINARY: PRT: 456 AA.
AC O23270;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)
DE GLUCOSYLTRANSFERASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98121113.
RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,
RA BERKAMP R., DIRKSE W., VAN STAVEREN M., STIEKEMA W., DROST L.,
RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFANELLI P., WEDLER H.,
RA WEDLER E., WAMBUIT R., WEITZENEGGER T., POHL T.M., TERRIN N.,
RA GIELEN J., VILLARROEL R., DE CLERCK R., VAN MONTAGU M., LECHARNY A.,
RA AUBORG S., GY I., KREIS M., LAC N., KAVANAGH T., HEMPEL S., KOTTER P.,
RA ENTIAN K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,
RA SILVEY M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,
RA VOURELATOU E., MILIONI D., HATZIOPOULOS P., PIRAVANDI E., OBERMAIER B.,
RA HILBERT H., DUESTERHOFT A., MOORES T., JONES J.D.G., ENEVA T.,
RA PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,
RA DELENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,
RA SCHUELLER C., CHALWATZIS N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana.";
RT Nature 391:485-488(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 297335; CAB10189.1;
DR PROSITE; PS00375; UDRGT; 1.
DR PFW; PF00201; UDRGT; 1.
DR Transferrase.
SQ SEQUENCE 456 AA; 50702 MW; EAE0D380 CRC32;

Query Match 44.1%; Score 1090.5; DB 10; Length 456;
Best Local Similarity 47.3%; Pred. No. 1.1e-71;
Matches 221; Conservative 85; Mismatches 132; Indels 29; Gaps 9;

QY 3 QPHVILTPFAQHINPALQFAKNLVKMGIEVTFSTSIYAQSRMDEKSLNAPKGLNFTP 62
Db 11 RPHLVTPFAQHINPALQANLRIHGAATVTVTAHRRMGPP---STKGLSPAW 67
QY 63 FSDGDEGFDHSDKDPVYMSQLKCGSETVKII---LPCSENGOPTCLLYSIFLPWAA 119
Db 68 FTDGDDGLKSFEDQKIYNSLKRCSNALRDIKANLADATTETPITGVISVLVPWYS 127
QY 120 EVAREVHIPSALLWSQPATILDIYFNFHGYEKAMANESNDPNWISQLPGLP 179
Db 128 TVAREHPLTLLIWPATVLDIYFYFNTSYKHLFDVE-----PIKPKPLITGDL 182
QY 180 SLLPYGANGSLRVALPPPKELIDTLDATTPKILVNTFDELEPEALNAIEGKFGIGP 239
Db 183 SLQPSKA---LPSALVTUREHIALETESNPKILVNTFSALEHDALTSVEKLMPIGP 239
QY 240 LIPSAPLGGNDPLDASFGGDLFQNSN-DYMEWLNKSPNSVVIYSPGSLMN--PSISOME 296
Db 240 LVSS-----SGKTDLFKSSDEDTYKWLDSKLSRVIIISLCTHADDLPE-KHME 288
QY 297 ETSKGLIDIGRFLVMYIKENEGKEENKKGICIELEKIGKIVPWCQSLYKHPISLGC 356
Db 289 ALTHGVLATNRPLWIVREKNPEEKKRNFLEIRGSDR-GLVVGWCSQTAVLHCAVGC 347

QY 357 FYSHGWNLSALESLAGVPVAFPOWTDQMTNAKQVEDYWKSGVRVRIWEDGVSEETK 416
Db 348 FVTHGWNSTLESLSGVPVAFPOFADQCTTAKLVEDTWIRGVKVKVGEEDVDGEIR 407
QY 417 RCIELVMDGGEKEELRNKAKKELAREAVKEGGSSHKLNKAFID 463
Db 408 RCLEKVMGGEEAEEMRENAEKWAMAVDAAAEAGGSDLNKGFVDE 454

RESULT 7
Q92VY2 PRELIMINARY: PRT: 519 AA.
AC Q92VY2;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)
DE T25N20.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA SHINN P., BUEHLER E., DEMAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OUI O., SHEN Y.K., TORIUMI M.,
RA VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T25N20.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005106; AAC80599.1;
DR PROSITE; PS00375; UDRGT; 1.
DR SEQUENCE 519 AA; 58584 MW; C982007F CRC32;

Query Match 43.8%; Score 1084.5; DB 10; Length 519;
Best Local Similarity 48.2%; Pred. No. 3.6e-71;
Matches 232; Conservative 71; Mismatches 131; Indels 47; Gaps 14;

QY 1 MVQPHVILTPFAQHINPALQFAKNLVK-MGIEVTFST--SIYAQSRMDEKSLNAPKG 57
Db 51 MAPPHFLVTPFAQHINPALQFAKNLVKIRKTGARVTVTCVSVFHNMTANH---NKVEN 107
QY 58 LNFIPFSDGFDGSG---FDHSDKDPVYMSQLKCGSETVKIILTCSSENGOPTCLLYS 112
Db 108 LSELFSDGFDGSGISTYEDRQKRSV---NLKVGDKALSDFIETKMGDSPTVTLIT 163
QY 113 IFLPWAAEVAREVHIPSALLWSQPATILDIYFNFHGYEKAMANESNDPNWISQLPGL 172
Db 164 ILLNMAPKVARFQLPSALLWLPALVFNYYTHF-----MGKNS-----VPELPNLS 212
QY 173 LETRDLPSFLPYGA-KGSLRVALPPPKELIDTLDATTTPKILVNTFDELEPEALNAIEG 231
Db 213 LEIRDLPSFLTPSNTKNG---AYDAFQSMMEFLIKETPKILINTFDSLEALTAFTPN 268
QY 232 YKFGYGLPILPSAFGLGNDPLDASFGGDLFQNSNDYMEWLNKSPNSVVIYSPGSLMNS 291
Db 269 IDWAVGVLPLTPEIFSG-----STNKSVDQSSSYTLWLDKSTESSVIYVSGTMEVLS 322
QY 292 ISQMEISKGLDIDIGRPFLLWVI--KENEGK---EE---NKKLGICIELEKIGKIVPWC 343
Db 323 KQIEELARALLEGKRPFLWITDKSNRTKTEGEETEIEKIAGRHELEEVGLVSWC 382
QY 344 SQLEVLKHPSLGCFVSHCGWNSALESACGVVPVAFPOWTDQMTNAKQVEDYWKSGVRV 403
Db 383 SQIEVLSHRAVGCFTVHCWSSTLESVLGVVPVAFPMWSDQPTNAKLEESWKTGVRV 442
QY 404 INEDGVSEETKRCIELVMDGGEKEELRNKAKKELAREAVKEGGSSHKLNKAFID 463
Db 443 ENKDGILVERGEIRRCLEAYME--EKSVELRENAKKWRLAMEAGREGGSSDKNMEAFVD 500
QY 464 V 464
```

```

Db 501 i 501
RESULT 8
Q92VY5 PRELIMINARY; PRT; 455 AA.
AC Q92VY5;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
T25N20.17.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA SHINN P., BUHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORTUMI M.,
RA VYSOTSKAYA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RT *Genomic sequence for Arabidopsis thaliana BAC T25N20.*;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005106; AAC80596.1;
DR PROSITE; PS00375; UDPGT; 1.
SQ SEQUENCE 455 AA; 51190 MW; F9BAC33E CRC32;

Query Match 42.8%; Score 1060; DB 10; Length 455;
Best Local Similarity 46.9%; Pred. No. 1.8e-69;
Matches 223; Conservative 78; Mismatches 136; Indels 38; Gaps 12;

QY 1 MVQPHVLTTPAAGHINPALQFANKLVK--MGIEVTFST--SIYAQSRMDEKSIINAPK 57
Db 1 MAQPHLVLTTPAAGHINPALQFANKLVK--MGIEVTFST--SIYAQSRMDEKSIINAPK 57
QY 58 LNFIFDGFDFGPHSDPV--FYNSQLKCGSEVVKIILTCSENGOPTICLLYSIPL 116
Db 58 LSLFTDGFDFGPHSDPV--FYNSQLKCGSEVVKIILTCSENGOPTICLLYSIPL 117
QY 117 MAEVAEAREVHIPSALLWSQATILDIYVFNPHGYEKAMANESNDPNWSIQLPGLPLETR 176
Db 117 WPKVARHPLSVHLNTPQAFADIIY-----NYSTGNSVFEPNLPISLR 166
QY 177 DLPSFLPYGAGKSLRVALPPFKELIDTLDATTPKILVNTFDELEPALNAIEGYKFG 236
Db 177 DLPSFLSP---SNTKAAQAVYQELMDFLKEENPKILVNTFDSLEPEFLTAIPNIEVA 223
QY 237 IGPLPSAFLGNDPLDASFGDLFQ--SNDYMELNKSNSVVIYISFGSLMNPISQ 294
Db 224 VGPLPAEIFTGES-----GKDLSDHQSSYTLWLDSKTESVIYISFGTWVLSKKQ 278
QY 295 MEEISKGLDIGRPLWVI--KENEGK---EEE---NKKLGCIEELEKIGKIVPWCSQL 346
Db 279 IEELARALIEGRPLWITDKLNREAKEIEGEEFTEIEKIAFRLEEVGVIVSCSQI 338
QY 347 EVLKHPSLGCFVSHCGWNSALESACGVVPVAFPOWTDQMTNAKQVEDVWKGVRINE 406
Db 339 EVLRHRAIGCFTHCGWSSLSLVGVPVAFPMKSDQANAKLLEEITWKTGVRVENS 398
QY 407 DGVVESEETKRCIELVMDGEGEELRNKAKWKLAREAVKEGSSHNKAFI 461
Db 399 EGLVERGEITMRCLAEVMEA--KSVLELRENAEKWKRLATEAGREGGSDKNVEAF 451

RESULT 9
Q23406 PRELIMINARY; PRT; 458 AA.
AC Q23406;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

```

```

DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE SIMILAR TO INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA BEVAN M., STIEKEMA W., MURPHY G., WAMBUITT R., POHL T., TERRY N.,
RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
RA PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFF A.,
RA JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,
RA SCHUELLER C., CHALWATZIS N.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97339; CAB10333.1;
DR MENDEL; 26727; Arath; 2543; 26727.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 458 AA; 52005 MW; 67943A2F CRC32;

Query Match 42.4%; Score 1049.5; DB 10; Length 458;
Best Local Similarity 46.5%; Pred. No. 1.1e-68;
Matches 226; Conservative 78; Mismatches 115; Indels 67; Gaps 15;

QY 4 PHVLTTPAAGHINPALQFANKLVK--MGIEVTFSTSIYAQSRMDEKSIINAPKGLNFI 61
Db 12 PHFLVTPAAGHINPSLEAKRLAGTISGARVTFASISAYNRR--MFSTENVPETLIFA 70
QY 62 PFSDFGDFG-----DHKSDPVF--YNSQLKCGSEVVKIILTCSENGOPTICLLYSI 113
Db 71 TYSGDHGDGFKSSAYSDKSQDATGNFMSEMRRRGKETLTELIEDNRKQNRPFVCVYTI 130
QY 114 FLPAEVAEAREVHIPSALLWSQATILDIYVFNPHGYEKAMANESNDPNWSIQLPGLPL 173
Db 131 LLTWAELEA-----LFSIFYHYFNGFEDAISEMANTPSSSIKLPISLPL 174
QY 174 ETRDLPFLPYGAGKSLRVALPPFKELIDTLDATTPKILVNTFDELEPALNAI--EGY 232
Db 175 TVRDIPSEIVSNVYAF--LPAFREQIDSLKEEINPKILINTFQELEPEAMSSVPDNF 231
QY 233 KFYGIGLIPSAFLGNDPLDASTGGDLFQNSNDYMELNKSNSVVIYISFGSLMNPIS 292
Db 232 KIVPVGPLL-----TLRTD--FSSRGYIEIWDTKADSSVLYVSFGTLAVLSK 277
QY 293 SOMEIEISKGLDIGRPLWVI-----KENEKKEEENKKLGCI-----EELKIGKIVP 341
Db 278 KQVLELCALIQSRRPLWITDKSYRNKEDEKEED-----CISSEFRELDELGVVS 332
QY 342 WCSQLEVLKHPSLGCFVSHCGWNSALESACGVVPVAFPOWTDQMTNAKQVEDVWKGVR 401
Db 333 WCDQFRLNHRISGCFVTHCGWNSTLESIVSGVPVAFPOWQNDQMMNAKLEDCWKTCVR 392
QY 402 V--RINEDG---VWSEETKRCIELVMDGEGEELRNKAKWKLAREAVKEGSSHNK 457
Db 393 VMEKEEGGVVVDSEETRRRCIEEYME--DKAEFRGNATRWKDLAAEAVREGSSFNHL 450
QY 458 KAFIDD 463
Db 451 KAFVDE 456

RESULT 10
Q23401 PRELIMINARY; PRT; 479 AA.
AC Q23401;
DT 01-JAN-1998 (TREMBlrel. 05, Created)

```



OC Arabidopsis.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BEVAN M., STIEKEMA W., MURPHY G., WAMBUIT R., POHL T., TERRY N.,  
 RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,  
 RA PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFT A.,  
 RA JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,  
 RA SCHUELLER C., CHALWATZIS N.,  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z97339; CAB10328.1;  
 DR PROSITE: PS00375; UDCPT; 1.  
 DR PFAM: PF00201; UDCPT; 1.  
 KW Transferase.  
 SQ SEQUENCE 475 AA; 53495 MW; A87ECA5 CRC32;

Query Match 30.9%; Score 766; DB 10; Length 475;  
 Best Local Similarity 36.9%; Pred. No. 5e-48;  
 Matches 175; Conservative 91; Mismatches 172; Indels 36; Gaps 14;

QY 4 PHVLTTPAAGHINPALQFAKNLVKMGIEVTFSTSIYQAQRMDEKSIINAPKG 57  
 DB 8 PHVLTTPAAGHINPALQFAKNLVKMGIEVTFSTSIYQAQRMDEKSIINAPKG 57  
 QY 58 LNFPTSGDGDHSDKDPVFMQSKGSETVKKIILTCSENGQITCLLYSIFLPW 117  
 DB 67 LGFLRF-EFFEDGVPYKEDFDLLQKSLVSGKREIKNLVKYK- -QPVRCILINNAEVPW 123  
 QY 118 AAEVAREVHIPALLWSQPATILDIYVFNHGYEKAMANESNDPNWISQIPL-LETR 176  
 DB 124 VCDIAELQIPSAVLWQSCACLAAYYYHQLVKFTT-ETEPEITVDFPFRPLTKHD 181  
 QY 177 DLPSFLPYGAGSL-RVALPPFKELDTLDAETTPKILVNTFDELEPEALNAIE- -G 231  
 DB 182 EIPFLPSPSSIGTILFOIKRLKHP- -SVLIETFOELEKDIIDHMSQLCPQ 235  
 QY 232 YKFGIGLPLPSAFGLGNDPLDASFGDLFQNSNDYEWLNKSPSSVYISFGLSNPS 291  
 DB 236 VNFNPGLPTMA- -KTIRSDIKGDISKPDSDCIWLDSDREPSSVYISFGLTAPLK 290  
 QY 292 ISQMEESKGLIDIGRFLWIKENEKKEENKKGICIELEKIGKIVPCSOLEVLKH 351  
 DB 291 QNQIDELAHGLNLSGLSCLWVLRPPLGLELATEPHVLP- -ELEEKGIKIVWCQDEKVLAH 348  
 QY 352 PSLGCFVSHCGWNSALESACGVPVAFPOWTDQMTNAKQVEDYWKSGVRYR- -INEDGV 409  
 DB 349 PAVACFLSHCGWNSMEALTSVGVICFPQWQDQVTNAVYMDVFKTGLRLSRGASDERI 408  
 QY 410 VESEI-KRCIELVMDGEGEELRNKAKKWLAREAVKEGSSHKNLAFID 462  
 DB 409 VPREEVAERLEAV- -GEKAVELRENARRKWEAEASAVAYGTSERNFQBEVD 460

RESULT 13

ID O48676 PRELIMINARY; PRT; 460 AA.  
 AC O48676;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE F316.2 PROTEIN.  
 GN F316.2  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;  
 RA FEDERSPIEL N.A., PALM C.J., CONWAY A.B., KURTZ D.B., CONWAY A.R.,  
 RA AU M., ARAUJO R., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y.,  
 RA QUI O., OSBORNE B.I., SHINN P., SUN H., TORIUMI M., VYSOTSKAIA V.S.,  
 RA YU G., ECKER J., THEOLOGIS A., DAVIS R.W.,  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC002396; AAC00570.1;  
 DR MENDEL; 27887; Arath; 3145; 27887.  
 DR PROSITE: PS00375; UDCPT; 1.  
 DR PFAM: PF00201; UDCPT; 1.  
 SQ SEQUENCE 460 AA; 51002 MW; 02F08C19 CRC32;

Query Match 30.8%; Score 762; DB 10; Length 460;  
 Best Local Similarity 38.2%; Pred. No. 9.3e-48;  
 Matches 187; Conservative 78; Mismatches 163; Indels 62; Gaps 17;

QY 2 VQPHVLTTPAAGHINPALQFAKNLVKMGIEVTFSTSIYQAQRMDEKSIINAPKGLNFI 61  
 DB 8 VKGHVILPYVPGVGHLPVQFAKRLVSKNVKVIATTTVASSITTPS-----LSVE 60  
 QY 62 PFDGPD- - - - -EGFDHSDKDPVFMQSKGSETVKKIILTCSENGQITCLLYSIFL 115  
 DB 61 PISDGFDFIPGIPGF- -SVDT- - -YSEFSLNGSETTLTLLIEKFKSTDSPDCLYDSFL 116  
 QY 116 PAAEVAAREVHIPALLWSQPATILDIYVFNHGYEKAMANESN- - - - -DPNWS-IOL 167  
 DB 117 PWGLEVARSHLSNAFFTNNLTVCV- - - - -LRKFSNGDFPLPADPNSAPFRI 165  
 QY 168 PGLPLLTRELPLSP- - - - -LPYGAKGSLRVALPPFKELDTLDAETTPKILVNTFDELEP 223  
 DB 166 RGLPSLSYDELPSVGRHLLTHPEHG- -RVLLNQFP- - - - -NHENADWLFPVNGFEGLE- 216  
 QY 224 EALNAISG- - - - -YKFGIGLPLPSAFGLGNDPLDASFGDLFQNSNDYEWLNKSPSS 278  
 DB 217 ETQDCENGESDAMKATLIGMIPSAIYLDORMEDDKYGASLLKPKISKECMEWLETKQAQS 276  
 QY 279 VYISFGLSNPSISOMEETSKGLIDIGRFLWIKENEKKEENKKGICIELEKIGK 338  
 DB 277 VAFVSGSFGILPEKGLAEVAIAEQESDLNPLWIKAKIAKLPE- - - - -GVESTKDRAL 332  
 QY 339 IVPWCSOLEVLKHPSLGCFVSHCGWNSALESACGVPVAFPOWTDQMTNAKQVEDYWK 398  
 DB 333 LVSCNCLVLAHESIGCFELTHCGWNSSTLEGLSGVPMGVQWSDQMDNAKFEVEYWKV 392  
 QY 399 GVRVRIEDG- - -VESEIKRCIELVMDGEGEELRNKAKKWLAREAVKEGSSHKN 456  
 DB 393 GYRAK-EEAGEVIVKSEELVRCLEKGVME-GESSVYKIRESSKKKDLAVKAMSEGGSDRS 450  
 QY 457 LKAFIDVAK 466  
 DB 451 INEFIESLGK 460

RESULT 14

O23400 PRELIMINARY; PRT; 484 AA.  
 ID O23400  
 AC O23400;  
 DT 01-JAN-1998 (Tremblrel. 05, Created)  
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BEVAN M., STIEKEMA W., MURPHY G., WAMBUIT R., POHL T., TERRY N.,  
 RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,  
 RA PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFT A.,  
 RA JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:51:36 ; Search time 59.57 Seconds  
(without alignments)  
120.431 Million cell updates/sec

Title: US-09-147-955-12

Perfect score: 2475

Sequence: 1 MVQPHVILTFPAQGHINPA.....EGGSSHKLNKAFIDVAKGF 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_AA:\*
- 1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/iaa/6.COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/iaa/PCRUS.COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	750.5	30.3	471	2	US-08-466-583-2
2	750.5	30.3	471	4	PCT-US95-07820-2
3	481	19.4	471	3	/cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
4	443.5	17.9	488	2	US-09-106-464-2
5	359	14.5	131	2	US-08-797-226-2
6	359	14.5	131	4	US-08-466-583-5
7	266	10.7	131	2	PCT-US95-07820-5
8	266	10.7	131	2	US-08-466-583-6
9	225	9.1	63	2	PCT-US95-07820-6
10	225	9.1	63	4	US-08-466-583-8
11	189	7.6	531	4	PCT-US95-07820-8
12	183	7.4	534	4	PCT-US92-00282-6
13	173	7.0	529	4	PCT-US92-00282-4
14	173	7.0	533	4	PCT-US92-00282-7
15	168	6.8	531	4	PCT-US92-00282-3
16	128.5	5.2	506	5	PCT-US92-00282-5
17	123	5.0	56	2	5180581-2
18	123	5.0	56	2	US-08-466-583-4
19	114	4.6	74	4	PCT-US95-07820-4
20	112	4.5	409	2	PCT-US92-00282-24
21	112	4.5	409	3	US-08-924-254-2
22	107.5	4.3	414	2	US-09-120-249-2
23	106	4.3	3443	2	US-08-750-524-1
24	104.5	4.2	510	1	US-08-416-603-2
25	102.5	4.1	511	3	US-08-278-635B-4
26	100	4.0	58	2	US-08-464-258B-4
27	100	4.0	58	4	US-08-466-583-9
28	98	4.0	52	2	PCT-US95-07820-9
					US-08-466-583-7

29	98	4.0	52	4	PCT-US95-07820-7	Sequence 7, Appli
30	97.5	3.9	529	1	US-08-496-855A-2	Sequence 2, Appli
31	96	3.9	528	2	US-08-466-589-2	Sequence 2, Appli
32	96	3.9	528	2	US-08-700-636-2	Sequence 2, Appli
33	96	3.9	528	3	US-08-467-574-2	Sequence 2, Appli
34	94	3.8	780	2	US-09-018-760-4	Sequence 4, Appli
35	92.5	3.7	2391	2	US-08-446-855A-2	Sequence 2, Appli
36	89.5	3.6	3031	1	US-07-689-008-2	Sequence 2, Appli
37	88	3.6	1220	2	US-08-843-530B-36	Sequence 36, Appli
38	85	3.4	775	2	US-08-966-388-4	Sequence 4, Appli
39	85	3.4	775	3	US-09-188-403-4	Sequence 4, Appli
40	84	3.4	1022	1	US-08-271-364A-8	Sequence 8, Appli
41	84	3.4	1022	2	US-08-222-715B-27	Sequence 27, Appli
42	83.5	3.4	408	2	US-08-926-258-2	Sequence 2, Appli
43	83.5	3.4	408	2	US-09-120-053-2	Sequence 2, Appli
44	83.5	3.4	552	3	US-09-111-752-5	Sequence 5, Appli
45	83.5	3.4	897	1	US-07-960-389-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-08-466-583-2  
; Sequence 2, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szczygowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: and Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,583  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 471 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-466-583-2

Query Match 30.3%; Score 750.5; DB 2; Length 471;  
Best Local Similarity 35.5%; Pred. No. 1.2e-68;  
Matches 172; Conservative 85; Mismatches 188; Indels 39; Gaps 13;





OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/106,464  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/055,554  
 FILING DATE: 13-AUG-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Spector, Eric S.  
 REGISTRATION NUMBER: 22495  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-415-1500  
 TELEFAX: 703-415-1508  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 471 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-106-464-2

Query Match 19.4%; Score 481; DB 3; Length 471;  
 Best Local Similarity 28.9%; Pred. No. 6.3e-41;  
 Matches 153; Conservative 81; Mismatches 167; Indels 128; Gaps 22;

Qy 1 MVQPHVLTTPAGCHINPALQFANLVKMGIEVTFST-----SIYAQRMDK----- 49  
 Db 2 MGOLHFFFPMAQCHMPTLDMAKLVACRGVKATITPLNESVFSKAIRNKHGLTEI 61  
 Qy 50 --SILNAP-----KGLNFPFSGDFGDFGDKSDPVFMSQLRKCGSEVTKKIL 97  
 Db 62 DIRLLFPKAKENDPEDCERLDLPVSD-----KLPNFKAAAM--KDFEELIG 110  
 Qy 98 TCSENGOPITCLXSIPLPAAEVAAREVHIPSALLWSQPATILDIYFNFHG----- 149  
 Db 111 ECRPD-----CLVSDMFLPWTDSAAKFSIPRV-----FHGTSYFALCV 150  
 Qy 150 -----YEKAMANESNDPNSIQLEGLPLETRDLPFLPYGAGSLKVALPPKELIDT 204  
 Db 151 GDSIRRNKPFKNVSSD-----TETVVVPDF--PHEIR--LTRQLSPFEQ---- 191  
 Qy 205 LDAET--TPKI-----LVNFTFDELP---EALNAIEGKYKYGIGPLIPSFL 246  
 Db 192 SDEETGMAPMKAVRESDAKSYGFNFSELESYVEHYTKVGRKNWAIPL---SLC 248  
 Qy 247 GGNPDLDASFGDLFQNSNDYMEMLNKPNSVYVYISFGSLMNPISOMEISKGLIDIG 306  
 Db 249 NRDIYKAERKSSIDHACLKWLDSKSSIVYVCGSTADFTTAQMLANGLEASG 308  
 Qy 307 RPLWVIKE-NE-----KQKEENKKGICIELEKIGKIVPWCQSLVLELKHPSLCGFSHC 361  
 Db 309 QDFIWIIVRTGNEMLPEGEERTKEKGI-----IRGWAQVILLDHEAIGAFVTHC 360  
 Qy 362 GWSNALESAGVVPVAPQPTDQNTAKQVEDVVKSGVRV-----RINCDGVVSESEI 415  
 Db 361 GWNSTLEGISAGVPMLTWPVFAEQFNEKLVTEVNRSGAGVQSKWKQTASEG-VKREAI 419  
 Qy 416 KRCITLVMDGGEKGBELKAKNKAAREAVKEGSGSHKNAKAFIDV 464  
 Db 420 AKAIKRVMASEETEGFSRAKEKEMAREATEEGSGSINGWATLIQDI 467

RESULT 4  
 US-08-797-226-2  
 ; Sequence 2, Application US/08797226  
 ; Patent No. 5959180  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MOERS, CHARLES P  
 ; APPLICANT: ALLEN, PAUL V  
 ; APPLICANT: ROCKHOLD, DAVID R

APPLICANT: STAPLETON, ANDREW  
 APPLICANT: GARBARINO, JOAN E  
 APPLICANT: FRIEDMAN, MENDEL  
 APPLICANT: BELKNAP, WILLIAM R  
 TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE  
 TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE  
 TITLE OF INVENTION: GLYCOALKALOIDS IN SOLANACEOUS PLANTS  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: NANCY J. PARSONS  
 STREET: 800 BUCHANAN ST.  
 CITY: ALBANY  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94710  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA: US/08/797,226  
 FILING DATE:  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: PARSONS, NANCY J  
 REGISTRATION NUMBER: 40,364  
 REFERENCE/DOCKET NUMBER: 0011.97  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (510) 559-5731  
 TELEFAX: (510) 559-5777  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 488 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-797-226-2

Query Match 17.9%; Score 443.5; DB 2; Length 488;  
 Best Local Similarity 27.7%; Pred. No. 4.8e-37;  
 Matches 140; Conservative 95; Mismatches 191; Indels 79; Gaps 21;

Qy 5 HVILTTTFAQCHINPALQFANLVKMGIEVTFST-----IYAQRMDKSIILNAPKGLN 59  
 Db 12 HVLFPFLSAGHFIPLVNAARLAFASRGVKATILTPHALLFRSTIDDDVRISGPPISIV 71  
 Qy 60 FIPFSD---GFDEG---FDHSDKDP-----VFY-MSQLRKCGSETVKIILTCSENGQPT 107  
 Db 72 TIKFSAEVLPEGLIESFNSATSPENPHKIFVALSLQKPWEDKIREL-----RP-D 122  
 Qy 108 CLXSIPLPAAEVAAREVHIPSALLWSQPATILDIYFNFHGIEKAMANESNDPNSIQ 167  
 Db 123 CIFSDMYFPVTDIADELHIP--RIYLNLSAYMCYSIMHNLKVY--RPHKQPNLDESQSFV 180  
 Qy 168 PGLPLETRDLPFLPYGAGSLRVA---LPPFRELIDTLD--AETPKILVNTFDELE 222  
 Db 181 PGLP-----DEIKFKLSQ--LTDDLKSDQKTVFDELLEQVEDSESYGVYHDTFYELE 234  
 Qy 223 P---EALNAIEGKYKYGIGPLIPSALFGNLDPLDASFGDLFQNSND-YMEWLNKSPNS 278  
 Db 235 PAYDYYOKLKKPKCWHFGPLSHFASKIRSKELISE-----HNNNEIVIDLNAQPKS 288  
 Qy 279 VVYISFGSLMNPISOMEIEISKGLIDIGRPFLLWIKENEK-----GKEENKKGICIE 331  
 Db 289 VLYVSFGSMAREPESQNLNEIAQALDASNPFFIVLRPNEETASVLPVGNLEDKTKKGLY- 347  
 Qy 332 ELEKTKIVPWCQSLVLELKHPSLCGFSHCWSNALESACGVVPVAPQPTDQNTAKQ 391  
 Db 348 -----IKGWVPQLTIMEHSATGCFMTHCGTNSVLEAITFGVPMTWPLYADQFYNEKV 400  
 Qy 392 VE-----DWKSGVRVINEGDDGVVSESEIKRCIE--LYMDGGEKGEELRNKNAKW 439

Db 401 VEVRLGKIGIDVWNEGIEI---TGPVIESAKIREAIERLMISNGSEIINIRDRVMAM 457  
QY 440 KELAREAVKEGSSHNKAFIDDV 464  
Db 458 SRMAQNAETGEGSSWNLTAHQH 482

RESULT 5  
US-08-466-583-5  
; Sequence 5, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szerszen, Jędrzej B.  
; APPLICANT: Szczylowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: Genetic Control of Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,583  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 131 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: Internal  
US-08-466-583-5

Query Match 14.5%; Score 359; DB 2; Length 131;  
Best Local Similarity 48.9%; Pred. No. 2.7e-29;  
Matches 65; Conservative 26; Mismatches 36; Indels 6; Gaps 2;

QY 269 EWLNSKPNSVYISFGSLMNPISOMEETSKGLIDIGRPFVLWIKENEGKEENKKG 328  
Db 3 KWLDTFDRSVAVYVSGSLASLGNKAEELARGLLAAGKPFLLWV----RASDEHQVPY 58  
QY 329 CIELEKIG--KIVPMCSQLEVLKHPSLGCFVSHCGNNSALESACGVPVYAFPOMTDQM 386  
Db 59 LLAEATATGAAMVVPWCPQLDVLHAHPVGCFTVHCGWNSTLEALSFGVPMVAMALWTDQP 118  
QY 387 TNAKQVEDVWKS 399  
Db 119 TNARNVELAWG 131

US-08-466-583-5  
; Sequence 5, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szerszen, Jędrzej B.  
; APPLICANT: Szczylowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: Genetic Control of Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,583  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 131 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: Internal

RESULT 6  
PCT-US95-07820-5  
; Sequence 5, Application PC/TUS9507820  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szerszen, Jędrzej B.  
; APPLICANT: Szczylowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: Genetic Control of Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/07820  
; FILING DATE: 19-JUN-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94B PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 131 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: Internal  
PCT-US95-07820-5

Query Match 14.5%; Score 359; DB 4; Length 131;  
Best Local Similarity 48.9%; Pred. No. 2.7e-29;  
Matches 65; Conservative 26; Mismatches 36; Indels 6; Gaps 2;

QY 269 EWLNSKPNSVYISFGSLMNPISOMEETSKGLIDIGRPFVLWIKENEGKEENKKG 328  
Db 3 KWLDTFDRSVAVYVSGSLASLGNKAEELARGLLAAGKPFLLWV----RASDEHQVPY 58  
QY 329 CIELEKIG--KIVPMCSQLEVLKHPSLGCFVSHCGNNSALESACGVPVYAFPOMTDQM 386  
Db 59 LLAEATATGAAMVVPWCPQLDVLHAHPVGCFTVHCGWNSTLEALSFGVPMVAMALWTDQP 118  
QY 387 TNAKQVEDVWKS 399  
Db 119 TNARNVELAWG 131

US-08-466-583-6  
; Sequence 6, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szerszen, Jędrzej B.  
; APPLICANT: Szczylowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: Genetic Control of Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/07820  
; FILING DATE: 19-JUN-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94B PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 131 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: Internal  
PCT-US95-07820-5

APPLICANT: Bandurski, Robert S.  
 APPLICANT: Szczygowski, Krzysztof  
 TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
 TITLE OF INVENTION: and Plant Growth.  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee & Winner, P.C.  
 STREET: 5370 Manhattan Circle, Suite 201  
 CITY: Boulder  
 STATE: Colorado  
 COUNTRY: United States of America  
 ZIP: 80303

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/466,583  
 FILING DATE: 06-JUN-1995

CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/265,427  
 FILING DATE: 24-JUN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Ferber, Donna M.

REGISTRATION NUMBER: 33,878

REFERENCE/DOCKET NUMBER: 11-94A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 303/499-8080

TELEFAX: 303/499-8089

TELEX: 49617824

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 131 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

FRAGMENT TYPE: internal

US-08-466-583-6

Query Match 10.7%; Score 266; DB 2; Length 131;  
 Best Local Similarity 38.7%; Pred. No. 9.8e-20;  
 Matches 53; Conservative 22; Mismatches 50; Indels 12; Gaps 2;

Qy 268 MEWLNSKPNSSVYISFGSLMNPISOMEIISKGLIDIGRPFVWIKENE-----KGKEE 322

Db 2 LAWLGRQARGVAVSFGTVACPRDELRELAAAGLDSGAPFLMSLRDSDPHLPFGFLD 61

Qy 323 ENKLGCTEELEKIGKIVPWCSQLVLEKHPSLGCFVSHCGWNSALESACGVPVYAFQW 382

Db 62 RAAGTGS-----GLVVPWAPQVAVLRHPSVGAFVTHAGNASVLEGLSSGVPMACRPF 114

Qy 383 TDQMTNAKQVEDVWKS 399

Db 115 GDQRMNARSVAHVWGFG 131

# RESULT 8

PCT-US95-07820-6

Sequence 6, Application PC/TUS9507820

GENERAL INFORMATION:

APPLICANT: Bandurski, Robert S.

APPLICANT: Szczygowski, Krzysztof

TITLE OF INVENTION: Genetic Control of Plant Hormone Levels

TITLE OF INVENTION: and Plant Growth.

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee & Winner, P.C.  
 STREET: 5370 Manhattan Circle, Suite 201  
 CITY: Boulder  
 STATE: Colorado  
 COUNTRY: United States of America  
 ZIP: 80303  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/07820  
 FILING DATE: 19-JUN-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/265,427

FILING DATE: 24-JUN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Ferber, Donna M.

REGISTRATION NUMBER: 33,878

REFERENCE/DOCKET NUMBER: 11-94B PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 303/499-8080

TELEFAX: 303/499-8089

TELEX: 49617824

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 131 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

FRAGMENT TYPE: internal

PCT-US95-07820-6

Query Match 10.7%; Score 266; DB 4; Length 131;  
 Best Local Similarity 38.7%; Pred. No. 9.8e-20;  
 Matches 53; Conservative 22; Mismatches 50; Indels 12; Gaps 2;

Qy 268 MEWLNSKPNSSVYISFGSLMNPISOMEIISKGLIDIGRPFVWIKENE-----KGKEE 322

Db 2 LAWLGRQARGVAVSFGTVACPRDELRELAAAGLDSGAPFLMSLRDSDPHLPFGFLD 61

Qy 323 ENKLGCTEELEKIGKIVPWCSQLVLEKHPSLGCFVSHCGWNSALESACGVPVYAFQW 382

Db 62 RAAGTGS-----GLVVPWAPQVAVLRHPSVGAFVTHAGNASVLEGLSSGVPMACRPF 114

Qy 383 TDQMTNAKQVEDVWKS 399

Db 115 GDQRMNARSVAHVWGFG 131

# RESULT 9

US-08-466-583-8

Sequence 8, Application US/08466583

Patent No. 5919998

GENERAL INFORMATION:

APPLICANT: Bandurski, Robert S.

APPLICANT: Szczygowski, Krzysztof

TITLE OF INVENTION: Genetic Control of Plant Hormone Levels

TITLE OF INVENTION: and Plant Growth.

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee & Winner, P.C.

STREET: 5370 Manhattan Circle, Suite 201

CITY: Boulder

STATE: Colorado

COUNTRY: United States of America

ZIP: 80303





; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 529 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US92-00282-7

Query Match 7.0%; Score 173; DB 4; Length 529;  
Best Local Similarity 22.2%; Pred. No. 3.7e-09;  
Matches 98; Conservative 57; Mismatches 149; Indels 138; Gaps 19;

QY 91 TVKKIILTCSENGQPTCLLYSIFL-----PWAAEVAREVH----- 126  
DB 40 SKREIVHLSERGHDIIVLPEVNNLLGSKYYRRKSPFPYNNLELRTRYRSFGNNHFA 99  
QY 127 IPSALL-----WSQATILDIYFNFHGYEKAMANES-----NDPNWSIQLP-G 169  
DB 100 ASSPLMAPLREYNNMIVDMCFSCQSLKDSATLSFLRENQFDALFTDP-----AMPQG 155  
QY 170 LPLETRDLPFLPYGAGSLR-----VALPPFKELIDTLDATTPPKILVNTFDE 220  
DB 156 VILAEYKLPFIYLFGRFPCSLHIGQSPSPVSVYPRFYTKFS--DHMTFPQRLANFI-- 211  
QY 221 LEPEALNATEGKFFYGI-----GFLIP 242  
DB 212 -----ANILENTLYHCLYSKEYILASDLKRDVSLPALHQNLSMLLRYDFVFPYPRVMP 266  
QY 243 S-AFLGNDPLDASFGDILFQNSNDYMWLNKPNSSVYVYISFGSLMN--PSISOMEELS 299  
DB 267 NMIFIGT---NCKKKNL---SQEFAVYNASGEHGIYVFSLGSMVSEIPEKKAM-EIA 319  
QY 300 KGLIDIGRFLVYKNEKGKEENKKGICIELEKIGKIVPCWOLEVLPKHPSLGCFVYS 359  
DB 320 EALGRIPQTLW---RYTGTRPSN-----LAKNTILVKWLPPNDLLGHPKARAFIT 367  
QY 360 HCGWNSALESACGVVPAFPQWTDQMTNAKQVEDYWKSGVRRVINEGDVYVESEELKRCI 419  
DB 368 HSGSHIYEGICNGVPMVMPPLFGDQMDNAKME---TRGAGVTLN-----VL 412  
QY 420 ELVMDGGEELRKNNAKWKKE 441  
DB 413 EXTADLENALKTVNNKSYKE 434

RESULT 14  
PCT-US92-00282-3  
; Sequence 3, Application PC/TUS9200282  
; GENERAL INFORMATION:  
; APPLICANT: OWENS, IDA S.  
; APPLICANT: RITTER, JOSEPH K.  
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION  
; TITLE OF INVENTION: THEREIN.  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
; STREET: 1615 L STREET, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00282  
; FILING DATE: 19920110  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26581  
; REFERENCE/DOCKET NUMBER: 91532-PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-861-3000  
; TELEFAX: 202-822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 533 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US92-00282-3

Query Match 7.0%; Score 173; DB 4; Length 533;  
Best Local Similarity 23.1%; Pred. No. 3.7e-09;  
Matches 87; Conservative 64; Mismatches 126; Indels 100; Gaps 19;

QY 93 KKIILTCSENGQPTCLLYSIFLPHAAEVAREVHIPSALLWSQATILDIYFNFHGYEK 152  
DB 134 KELMASLAESSFDV--MLTDPFLPCSPIVAQVLSLPTV-----FFLHALPC 177  
QY 153 AMANESND--PNWSIQLPGLPLETRDLPF-----LLPYGAKGSLRVALPPEKELI-D 203  
DB 178 SLEFATQCPNFSYVP-RPLSSHSDHMTFLORVKNKLAFSONFLCDVYVYSPYATLASE 236  
QY 204 TLDAETTPKILVNT-----FDELEPEALNAIEGYKFGYIGPLIPS-AFLGNDPLD 253  
DB 237 FLQREVTVQDLLSSASVWLFSDFKVDYR-----PIMPNVVFVGIN--- 279  
QY 254 ASFGGDLFON--SNDYMWLNKPNSSVYVYISFGSLMNPISOMER-----ISKGLIDIG 306  
DB 280 -----CLHONPLSQEPEAYINASGEHGIYVFSLGSM---VSEIPEKKAMATADALGNP 330  
QY 307 RPLVAVIKENEKGKEENKKGICIELEKIGKIVPCWOLEVLPKHPSLGCFVSHCGWNSA 366  
DB 331 QTVLW---RYTGTRPSNLANNTI-----LVKLPNDLLGHPWTRAFITHAGSHGV 378  
QY 367 LESLACGVVPAFPQWTDQMTNAKQVEDYWKSGVRRVINEGDVYVE--SSEIKRCIELVMD 424  
DB 379 YESICNGVPMVMPPLFGDQMDNAKME---TKGAGVTLN---VLENTSEDLLENALKAVI- 431  
QY 425 GGEKGEELRKNNAKWKKE 441  
DB 432 -----NDKSYKE 438

RESULT 15  
PCT-US92-00282-5  
; Sequence 5, Application PC/TUS9200282  
; GENERAL INFORMATION:  
; APPLICANT: OWENS, IDA S.  
; APPLICANT: RITTER, JOSEPH K.  
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION  
; TITLE OF INVENTION: THEREIN.  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
; STREET: 1615 L STREET, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00282

```

; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-00282-5

```

```

Query Match      6.88; Score 168; DB 4; Length 531;
Best Local Similarity 20.08; Pred. No. 1.2e-08;
Matches 92; Conservative 67; Mismatches 134; Indels 166; Gaps 20;

Qy 19 PALQAKNLVKMGIEVTFSTSIYAQSRMDEKSIILNAPK-----56
Db 108 POTEYRNMIIVIGLYF-----INQSLLODRDTLNFKESKFDALFTDPALPCGVILAEY 162
Qy 57 -GLNFIPIFDGTF---DEGPDHSDVPFYMSQLRKCGSETVKKIILTCSENGQDITCLLY 111
Db 163 LGLPSVYLFRGFCPSLEHTFSRSPDPVSY---IPRCYTKFSDHMTFS---QRVANFLV 214
Qy 112 SIFLPMAEVAEVAREVHIFSAIWSQPATILDIYFNFHCYKAMANESNDPNWSIQLPLP 171
Db 215 NLEPI-----LFYCLFSKYK-----231
Qy 172 LLETRLPSFLLPYGAKGSLRVALPFPKELDTLD-AETTPKILVNTFDELEPEALNAIE 230
Db 232 -----LASAVLKRVDVITITLSEVSVMLLRDYDFVLEYP-----264
Qy 231 GYKFGYIGLIPS-AFLGGNDPLDASFGGDLFQNSNDYMEWLNKSPSSVYVIFSGSLMN 289
Db 265 -----PVMNPMVFVIGG---INCKRKKDL---SQEFAYINASGEHGIYVFSLSGM-- 308
Qy 290 PSISQMEE-----ISKGLIDIGRPFVWIKENEGKKEEENKKGCIIELEKIGKIVPWS 344
Db 309 --VSEIPEKKAMATADALGKNPQVLM-----RYTGTSPSLNANTTI-----LVKWL 354
Qy 345 QLEVLKHPSLGCFVSHCGWNSALESACGVPVAFPPQWTDQMTNAKOVEDYWKSGVRVRI 404
Db 355 QNDLLGHMPTAFITHAGSHGVYESICNGVPMVMPVLPFGDQMDNAKME---TKGAGVTL 411
Qy 405 NEDGVVE--SEEIKRCIELVNDGGEKEELRKNKAKWKE 441
Db 412 N---VLEMTSEDLENALKAVI-----NDKSYKE 436

```

Search completed: August 1, 2000, 19:51:38  
Job time: 18760 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:17:13 ; Search time 152.01 Seconds  
(without alignments)  
1512.110 Million cell updates/sec

Title: US-09-147-955-7

Perfect score: 1671

Sequence: 1 aacacataaaaaataa.....ctaaaaaataaaaaaa 1671

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*
- 7: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101.8	6.1	1731	4	US-08-466-583-1
2	101.8	6.1	1731	6	PCT-US95-07820-1
3	62.6	3.7	7218	1	US-08-232-463-14
4	48.6	2.9	1669	3	US-08-522-421-1
5	45.4	2.7	1607	4	US-08-797-226-1
6	45.4	2.7	1627	5	US-09-106-464-1
7	40	2.4	405	4	US-08-975-316-32
8	38.8	2.3	4615	1	US-08-188-582-1
9	38.8	2.3	4615	1	US-08-646-715-1
10	38.4	2.3	1738	3	US-08-379-482A-2
11	37	2.2	1858	4	US-08-909-965C-11
12	37	2.2	53526	5	US-08-658-136-2
13	37	2.2	53577	5	US-08-658-136-2
14	36	2.2	251	3	US-08-623-906A-16
15	36	2.2	260	3	US-08-520-678A-29
16	36	2.2	2621	4	US-08-553-619B-8
17	35.6	2.1	806	5	US-08-906-769-120
18	35.6	2.1	1582	5	US-08-545-196B-10
19	35.6	2.1	1582	5	US-08-545-196B-12
20	35.4	2.1	855	7	5185441-40
21	35.4	2.1	855	7	5223394-3
22	35.4	2.1	863	1	US-07-940-861-11
23	35.4	2.1	863	2	US-08-459-512-11
24	35.4	2.1	863	4	US-08-459-657-11
25	35.4	2.1	863	4	US-08-460-132-11
26	35.4	2.1	863	6	PCT-US92-02050-11

27	35.4	2.1	863	7	5185441-35	Patent No. 5185441
28	35.4	2.1	863	7	5223394-5	Patent No. 5223394
29	35.4	2.1	1078	7	5223394-10	Patent No. 5223394
30	35.4	2.1	2280	1	US-08-320-151-11	Sequence 11, Appl
31	35.4	2.1	2280	1	US-08-413-118-11	Sequence 11, Appl
32	35.4	2.1	2280	5	US-08-473-446-11	Sequence 11, Appl
33	35.4	2.1	2993	4	US-08-415-593-42	Sequence 42, Appl
34	35.2	2.1	568	2	US-08-582-257-20	Sequence 20, Appl
35	35.2	2.1	568	3	US-08-582-298-20	Sequence 20, Appl
36	35.2	2.1	748	2	US-08-361-467B-3	Sequence 3, Appl
37	35.2	2.1	748	2	US-08-484-332C-3	Sequence 3, Appl
38	35	2.1	809	2	US-08-441-629-1	Sequence 1, Appl
39	35	2.1	809	5	US-08-776-207-1	Sequence 1, Appl
40	35	2.1	809	6	PCT-US95-03172-1	Sequence 1, Appl
41	34.8	2.1	796	1	US-08-104-073-2	Sequence 2, Appl
42	34.6	2.1	224	3	US-08-731-272A-26	Sequence 26, Appl
43	34.6	2.1	314	2	US-08-686-878A-25	Sequence 25, Appl
44	34.6	2.1	768	2	US-08-567-816A-1	Sequence 1, Appl
45	34.6	2.1	1841	6	PCT-US95-00362-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-08-466-583-1  
; Sequence 1, Application US/08466583  
; Patent No. 591998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szczyglowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: and Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,583  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,427  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 11-94A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1731 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 57..1472  
US-08-466-583-1



















**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:17:46 ; Search time 236.66 Seconds  
(without alignments)  
1766.547 Million cell updates/sec

Title: US-09-147-955-7  
Perfect score: 1671  
Sequence: 1 aacacataaaaaaaaaa.....ctaaaaaaaaaaaaaa 1671

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_36.:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1670.2	100.0	1671	1 X02829	WO9905287 Seq ID 4
2	484.4	29.0	2062	1 X02828	WO9905287 Seq ID 3
3	446.8	26.7	1474	1 X02827	WO9905287 Seq ID 2
4	434.8	26.0	1506	1 X02826	WO9905287 Seq ID 1
5	289	17.3	2105	1 X02831	WO9905287 Seq ID 6
6	103.6	6.2	568	1 T66174	UDP-glucose:thiohy
7	102	6.1	1513	1 T66166	UDP-glucose:thiohy
8	101.8	6.1	1731	1 T10106	Maize UDP-glucose
9	100.4	6.0	566	1 T66173	UDP-glucose:thiohy
10	83.8	5.0	432	1 T68693	Strawberry UDP-glu
11	78.4	4.7	1578	1 T62123	Solanum melongena
12	57.8	3.5	1597	1 V23108	UDP-glucose:flavon
13	48.4	2.9	1650	1 Q74684	Early Ripening Tom
14	45.4	2.7	1607	1 V49609	Potato solanidine
15	45.4	2.7	1624	1 V17054	Glucosyl transfera
16	45.4	2.7	1626	1 X24873	Medium-chain UDP-g
17	44	2.6	794	1 T66179	UDP-glucose:thiohy
18	42.8	2.6	940	1 T66178	UDP-glucose:thiohy
19	40.6	2.4	9789	1 T41852	CDNA encoding Plas
20	40	2.4	405	1 V23852	Plant CCR enzyme D
21	38.8	2.3	1486	1 X02855	Human zsig46 DNA.
22	38.8	2.3	4615	1 Q70724	TATA-binding prote
23	38.8	2.3	4615	1 T42210	Drosophila TATA-bi
24	38.8	2.3	4615	1 T79604	TATA-binding prote
25	38.4	2.3	1738	1 Q56241	Sequence encoding
26	38.2	2.3	2696	1 N60909	Plasmid pAU157 ins
27	38.2	2.3	2696	1 N60889	Plasmid pAU157 seq
28	38.2	2.3	2697	1 N60864	Sequence of plasm
29	38.2	2.3	2666	1 X32291	M. grisea PR2 gen
30	38.2	2.3	3383	1 T85161	Human chemokine re
31	38	2.3	9502	1 Q74770	Hepatitis C virus
32	37.8	2.3	223	1 T66184	UDP-glucose:thiohy
33	37.8	2.3	561	1 X20491	Human secreted pro

c 34 37.8 2.3 110000 1 V21209\_10 Continuation (11 o  
c 35 37.8 2.3 110000 1 V21209\_11 Continuation (12 o  
c 36 37.6 2.3 2401 1 X20301 Borrelia burgdorfe  
c 37 37.4 2.2 2418 1 Q27886 P.falciptarum GBPI3  
c 38 37.4 2.2 4696 1 X03041 Human IL-1ra BAC c  
c 39 37.4 2.2 5760 1 N50530 Sequence encoding  
c 40 37.4 2.2 6727 1 X02993 Human IL-1ra BAC c  
c 41 37.2 2.2 1302 1 X20437 Human secreted pro  
c 42 37 2.2 1858 1 T72175 Alzheimer's diseas  
c 43 37 2.2 53526 1 T94101 Human PKD1 gene. H  
c 44 37 2.2 53577 1 T18551 Human polycystic k  
c 45 37 2.2 53577 1 T94108 Human PKD1 locus b

## ALIGNMENTS

### RESULT 1

ID X02829 standard; cDNA; 1671 BP.  
AC X02829; 1999 (first entry)  
DE WO9905287 Seq ID 4.  
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.  
OS Torenia hybrida.  
FH Key Location/Qualifiers  
FT CDS 45..1481  
FT /tag= a  
FT /product= "protein with flavonoid 5-transglycosylation activity"

WO9905287-Al.

04-FEB-1999.

16-JUL-1998; J03199.

25-JUL-1997; JP-200571.

(SUNR ) SUNTORY LTD.

Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,

Yamazaki M;

WPI; 99-142940/12.

P-PSDB; W92950.

Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration

Disclosure; Page 60-64; 89pp; Japanese.

This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut flowers harvested from them.

Sequence 1671 BP; 530 A; 344 C; 399 G; 396 T;

Query Match 100.0%; Score 1670.2; DB 1; Length 1671;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1671; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacacataaaaaaaataaaagaaagaaataaataaaaaaaatgttaacaaacgcc 60

Db 1 AACACATAAAAAAAAAAATAAGAGAGAAATATTAATAAAAAAAAAAANTGTTAACAAAGCC 60

QY 61 atattctactagcaacattcccagcaagccacataaaacccctctctcgagttcgcca 120

Db 61 ATATTCTACTAGCAACATTCCAGCACAGGCCACATAAACCCCTTCTCTCGAGTTCCGCA 120

QY 121 aaaggctctcaacacccgagtagctgcacaaagtccattcttcacagtgatacagcat 180

Db 121 AAAGGCTCTCCAAACACCGGATAGCTGCACCAAGTCACATTTCTTCACGAGTGATACGCA 180

QY 181 tgaagcagcatgccttcgaaacccgagcagcagaatcgatttcgtgcatktycag 240

Db 181 TGAGAGCATGCGCTTCGAAACCGATCCGAGCAGCAGATGTTTCGTGGCATKTYCAG 240

QY 241 attcttacgatgatggcttaaaagaaagccagcagatggcaaaactacatgctcgagatga 300

Db 241 ATTCTTACGATGATGGCTTAAAGAAAGCCAGCATGTCGAGATGA 300

QY 301 gaaagcggaagcgaagccttaagacacactcttatttaagctcaagcagctcgatgg 360  
 DB 301 GAAAGCGGGAACGAAGCCCTTAAGGACACTCTTATTAAGCTCAACGATCGTCGATGG 360  
 QY 361 gaagtgaattacaacgcgctgagcttgggtgactctcactctatttttcgtggcgag 420  
 DB 361 GAAGTGAATGTTACAATCGCGTGAGCTTGGGTGTACTCTCATCTATTTTCGTGGCGAG 420  
 QY 421 ctgaagtggcgctgaagtcgagctgcgagtgccctcttcttttgatgaacccgctcagg 480  
 DB 421 CTGAAGTGGCGCTGAAGTCGAGCTGCGAGTGCCCTCTTCTTGGATTGAACCGCTACGG 480  
 QY 481 ttttcgattgtactatttttacttcaatgggtatgcgattgatgcgctggcctcag 540  
 DB 481 TTTTCGATGTACTATTTTACTTTCATGGGTATGCGATGATATCGATCGGGGCTCAG 540  
 QY 541 atcaaatccaactgcccaatcttcgcgagctctccaagcaagatccccctcttctctac 600  
 DB 541 ATCAAAATCCAATGCCCAATCTTCGCGAGCTCTCCAAGCAAGATCTCCCTCTTTCTCTAC 600  
 QY 601 tcccttcgagcccgagagcttcgaacccttaataagaagaaagtctcgacacgctcgaca 660  
 DB 601 TCCCTTCGAGCCCGGAGATTTCGAACCCCTTAATGAAGAAAGATTTCGACACGCTCGACA 660  
 QY 661 aagaccgaaagcgaagctcttgataaacacgcttcgacgcattagaaccgaaacaactca 720  
 DB 661 AAGAACCGAAAGCAAGGCTCTGTATAACAGCTTCGACGCAATTAGAAACCGCAACACTCA 720  
 QY 721 aagccatcgacaggtatgaactaatatccatcgggcccatatccatcatcgattctt 780  
 DB 721 AAGCCATCGACAGGTATGAACTAATATCCATCGGCCCATTAATCCCATCATCGATATTCT 780  
 QY 781 cagatggcaagcaccctcatcaagcaacaaatctctacggtggagacctcttcagaaaag 840  
 DB 781 CAGATGGCAACGACCCCTCATCAAGCAACAAATCTCTACGGTGGAGACCTCTTCAGAAAAG 840  
 QY 841 ccgatgaactacatagactggctgaactcaaacccggaactatcgctgttttacgcttt 900  
 DB 841 CCGATGAACCTACATGGACTGGCTGAACCTCAAAACCCGAATCATCGCTGCTTTACGTTT 900  
 QY 901 cgttcggagcctcctgaggtcccgaaaccccaaatggaagaataagcgaatagggcttt 960  
 DB 901 CGTTTCGGAGGCTCTCGAGGCTCCGAAACCCCAATGGAAGAAATAGCAATAGGCTTT 960  
 QY 961 cagacacaaatcgagcttctgggtgataagaagaacgaagagggcgacgaacaaag 1020  
 DB 961 CAGACACCAATCGCCAGTTCTCTGGGTGATPAGAAGAAACGAAGAGGGCGCACCAACAA 1020  
 QY 1021 agcaagcagaagaagaagagctgctgagctcttcttgatcgctcagggaaactgaacgac 1080  
 DB 1021 AGCAAGCAGAAGAGAGAGAGTGTGAGCTTCTTTGATCGTCACGGAACCTGAACGAC 1080  
 QY 1081 tcgggaaatcgtgacatgggtgcacaaattggatgttctgacgcataaagtcggtgggat 1140  
 DB 1081 TCGGGAAATCGTGACATGGTGTCTCACAATTTGGATGTCTCAGCATAAAGTCGGTGGGAT 1140  
 QY 1141 gcttcgagcgaatgggttggaaattctgctatcgagagcctggcttgggtgtgcccg 1200  
 DB 1141 GCTTCGAGCAGATGGGTTGGAAATTTGCTATTCGAGAGCGCTGGCTTGGTGTGGCCCG 1200  
 QY 1201 tgggtgcttccctcaatgggttcgatacaagggactaatgcgaagatgacgaagatggt 1260  
 DB 1201 TGGTGTGCTTCTCAATGGTTCGATCAAGGGGACTAATGCCAAGATGATCGAAGATGTGT 1260  
 QY 1261 ggaaggatgggtgagagtcagatgaaatgaggagcggcgttggatgagggcgtgaga 1320  
 DB 1261 GGAGGAGTGGTGTGAGAGTCAGATGAATGAGGAAGGGCGGCTTGTGTATAGGGGTGAGA 1320  
 QY 1321 ttaagaggtgcgtctcgaggttataaagactcgagagttgagagaaagcgaatgatgt 1380  
 DB 1321 TTAAGAGTGGCTCTCGAGGTTTAAAGAGTCGAGAGTTGAGAGAAAGCGCAATGATGT 1380

QY 1381 ggaaggggttgctaaagaagcctatggatgaagaaacgtggatcatcaatgaacaaactga 1440  
 DB 1381 GGAAGGGTTGGCTTAAGAAGCCTATGGATGAAGAACGTTGGATCATCAATGAACAACTGA 1440  
 QY 1441 agaatttttactagagattatttaataagctcctcataaagttgtactatattgttat 1500  
 DB 1441 AGAATTTTACTAGGATTATTAAATGAATGCTCATAGTTGTACTATATATGTTAT 1500  
 QY 1501 tattgttgatgaagcgtcgaattaaagattattagtttaaatgatattttagagaagaagc 1560  
 DB 1501 TATTGTGTTTATGACGCTCGAATTAAGTATTAGTTAAATGATATGTTATTTAGAGGAAGC 1560  
 QY 1561 caaaacgggctcacaccggcgagcggttgaaagcccgcatgatttataaata 1620  
 DB 1561 CAAAACGGGCTACACCGGCGAGCGGTTGGAAAGCCCGCATGATTTAAATATA 1620  
 QY 1621 tatttataataataatttttactattataactaaaaaataaaaaa 1671  
 DB 1621 TATTTAAATAAATATTTTCTACTATTAAACTAAAAAATAAAAAA 1671

RESULT 2  
 X02828  
 ID X02828 standard; cDNA; 2062 BP.  
 AC X02828;  
 DT 14-MAY-1999 (first entry)  
 DE W09905287 Seq ID 3  
 KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.  
 OS Verbena hybrida.  
 FH Key Location/Qualifiers  
 FT CDS 26..1411  
 FT /\*tag= a  
 FT /product= "protein with flavonoid 5-transglycosylation activity".  
 FT  
 PN W09905287-A1.  
 DR 04-FEB-1999.  
 PF 16-JUL-1998; J03199.  
 PR 25-JUL-1997; JP-200571.  
 RA (SUNR ) SUNTORY LTD.  
 PI Gong 2, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
 PI Yamazaki M;  
 DR WPI; 99-142940/12.  
 DR P-PSDB; W92949.  
 PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration  
 PS Disclosure: Page 58-60; 89pp; Japanese.  
 CC This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut flowers harvested from them.  
 CC  
 SQ Sequence 2062 BP; 608 A; 354 C; 508 G; 592 T;

Query Match 29.0%; Score 484.4; DB 1; Length 2062;  
 Best Local Similarity 63.4%; Pred. No. 2.3e-110;  
 Matches 920; Conservative 2; Mismatches 453; Indels 75; Gaps 9;

QY 31 aattataaaaaaaatggttaacaaacccatattctactagcaacattccagacacaag 90  
 DB 9 AAAAAAATAAAAAAATGAGCAGCTACGCTCTTGGCCACATATCCAGCAGCAGG 68  
 QY 91 gccacataaacctctctcgagttcgccaaaagcctcctcaaacggaatacgtcgacc 150  
 DB 69 GACACATAAATCCCGCCCTTCAATTCGCCAAGCGTCTCGCAATATGC-----CGACATTC 122  
 QY 151 aagtcacattctcacagagtgatatacgcattgagacgcagt---cgcttcgaaaccgcatc 207  
 DB 123 AAGTCACATTCTTCCACGCGCTTACGCGATGGCGCGCATGTCCAGAACCCGCGCTGCT 182  
 QY 208 cgagcagcagaatcgattctcgccatctkycagattcttaccagatgagcttaagaagaag 267  
 DB 183 CAAACGGGCTCATCAATTTTGTGCTTTTCCGACGGGTATGACGACGGGTACAGCCG 242

Qy	268	gcacgagtgcacaaactcaatctgttcggagatgagaaagcgcgggaacgaagccttaaaag	327
Db	243	GAGACGATGGGAAGAACTCATGTCTCGAGATGAAAGCAGAGAGGTATAAAAGCCTTGAGCG	302
Qy	328	acactcttattaagactcaacgatgcctgcgattgggaagtgaaattgttacaatcgcctgagct	387
Db	303	ATACTCTTGCAGCCAATATATGTCGATCAAAAAGACAGCAA-----ATCAGCT	350
Qy	388	ttgtggtactctatctatcttatttctgggcagctggaagtggcgtggaagtcgacatgc	447
Db	351	TCGTGGTGTACTCCACCTCTTTGCGATGGCGCGCAAGGTGGCGCGTGAGTTCCATCTCC	410
Qy	448	cgagtgcctcttcttgattgaacgcgtctacaggttttctgagtgtactatttttaactca	507
Db	411	GGAGCGGCTTACTCTCGATTGAGCAGCTACCGGTGTTGGATATATTTTACCTTTTATTTC	470
Qy	508	atgggtatgcctgatctcgatgcgggctcagatcacaatccaactgcgc---aatcttc	564
Db	471	ACGGCTATACGGACGAATCGATCGGGTTCGGATGCTATTTCACCTGCCGCGGAGCATCC	530
Qy	565	cgcagctctccaaagcaagatctcccctcttctactccccttcagcccccgcgagatcc	624
Db	531	CAGTGTGCGCCACGCGTGATTTACCGCTCTTCTCTTCTTCCATCCAGCATGAGAGATTC	590
Qy	625	gaaccctaatgaaagaaagtctcgacacgctcgacaaagaacccgaaagcgaagtccttga	684
Db	591	GTTCACTGATGAAGGAGAAATTGGAACCTTTAGAAGGTGAAGAAAACCTTAAGTCTTGG	650
Qy	685	taaacagcttcgacgcattagaataaccgaacactcaaaagccatcgacaggtatgaactaa	744
Db	651	TGAACACCTTTGATGCGTTGGAGCTGTATGCCCTCAAGGCCATTGATTAAGTACCGAGATG	710
Qy	745	tatccatcgggccattaatcccatcatcatcttcagatgycgaacgacccctcatcaa	804
Db	711	TTGCAATTGGCGCGCTTGATTCCTTCGCG---ATTCCTGGACGGTAAAGATCCTTCGGACA	767
Qy	805	qcaacaaatcctcagtgtagagacctcttcagaaaaagccgatgaaaaattcaatgagactggc	864
Db	768	GGTCTTTTGGCGGAGATTTGTCGAAAGGGTCTGAAATGACGACGATTCCTTCGAATGGT	827
Qy	865	taaactcaaaacccgaatcatcgcttcgtttacgcttctgctcgagagcctctcgagctcc	924
Db	828	TGACACAGTCTCGATCTTCGGTGGTTTACGTTTCGTTTCGTCGGAAGCTTCGTTAATCGA	887
Qy	925	cgaacccccaaatggaagaaatagcaatatgggcttcctagacaccacaaatcgccagttctct	984
Db	888	CGAAGTCGCAAAATGGAAGAGATAGCAAGAGGGCTGTAGATTGTGGGAGGCCGTTTTGT	947
Qy	985	gggtgataaagaagaacgaagggcgacgaacaagagcagaagcagaagaagaagaagc	1044
Db	948	GGGTGGTAAAGAGTAACGAAGGAGAAAGAGTA-----TTGATAGT	988
Qy	1045	tgctgagcttcttctgactcgcggaactgaacgactcgggaaatctgacatgggtgct	1104
Db	989	TGCATGGAGCAGTTGA-----AACGAGTGGGGAATAATTCATCTTGGTGT	1034
Qy	1105	cacaattggatgtcttgaacataaagtcggttgggatgtcttgtgacgcgattgcggttggga	1164
Db	1035	CTCAATTTGGAAGTCCGTGACGCATCCCTCGTGGGATGTTTCGTGTACACACTCGCGGTGGA	1094
Qy	1165	attctgctatcgagagcctggcttgggtgctcccgctggtgtgcttctcctcaatgggttcg	1224
Db	1095	ATTCCGACTTAGAGAGTATATCTTTCGGGGTTCGAGATGGCTTTTCCGAGTGGTTCG	1154
Qy	1225	atcaagggactaatgcgaagatgatacgaagatgtgtgagagagtggtgtgagatcagag	1284
Db	1155	ATCAAGGGACGAATGCGAAGCTGATGAGGATGCTGTGGAGCAGCGGTGTGAGATGAGAG	1214
Qy	1285	tgaatgagggaagcgcgttctgttgaatagcgttgagattaaaggtctgcgcgaggtta	1344
Db	1215	CTAATGAGAGGGTAGCGTGTGATGGTGAATTAAGAGATGTAATTCAGGAGGTTA	1274

Qy	1345	t-----aaagatcgagagttgagagaaagcgaatgatgtggaagggtttgg	1392
Db	1275	TGGATTGGGGGAGAAAGATTGAGAAACTTATGAGAGAGTCTGGCAAGTGGAAAGATTGG	1334
Qy	1393	ctaaagaagctatgatgaagaacgtggagatcatcaatgaacaatctgaagaattttat	1452
Db	1335	CAAGAAAAGCTATGGGGAAGA--TGGATCTCAGTTAACACCTCAAGGTCTTCTTG	1391
Qy	1453	ctaggattat	1462
Db	1392	ATGAGGTTGT	1401
RESULT	3		
X02827			
ID	X02827	standard; CDNA; 1474 BP.	
AC	X02827;		
DT	14-MAY-1999	(first entry)	
DE	W09005287	Seq ID 2.	
KW	Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.		
OS	Perilla frutescens.		
Key	Location/Qualifiers		
FH	29..1360		
FT	CDS		
FT	/tag= a		
FT	/product= "protein with flavonoid 5-transglycosylation		
FT	activity"		
PN	W09005287-Al.		
PD	04-FEB-1999.		
PF	16-JUL-1998; J03199.		
PR	25-JUL-1997; JP-200571.		
PA	(SUNR) SUNTORY LTD.		
PI	Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,		
PI	Yamazaki M;		
DR	WPI; 99-142940/12.		
DR	P-PSDB; W32948.		
PT	Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be		
PT	used to transform plants for improvement of plant coloration		
PS	Disclosure; Page 53-56; 89pp; Japanese.		
CC	This invention describes the isolation of plant proteins which have		
CC	flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used		
CC	in the production of plant varieties with improved colour, for example in		
CC	the production of flower varieties with desired coloration and cut		
CC	flowers harvested from them.		
SO	Sequence 1474 BP; 369 A;	348 C;	319 T;

Query Match	26.7%;	Score 446.8;	DB 1;	Length 1474;
Best Local Similarity	63.0%;	Pred. No. 3.8e-101;		
Matches 896;	Conservative 2;	Mismatches 434;	Indels 90;	Gaps 10;

  

QY	57	gcgcattcttactagcaacattccacgacaaaggccacataaaacctctctcaggttc	116
Db	38	CGCCGGGTGCTGTAGCAACGTTTCGGGCGAAGGCCACATAATCCGCCCTCCCAATC	97
QY	117	gccaaaaaggctctcacaacccggatcgtcgaccaagtccacattcttcagagtgtatac	176
Db	98	GCCAAAGACTCTCTAAAACCGG-----CACTGAGCTCACGTTTTTCAGAGGGTTTAT	151
QY	177	gcattgagcgcgatcgctctcgaaaccgatcogagcagcagaa-----tcgat	224
Db	152	GCATGGCGCGGTAGGCCAACACAGCTCCGGCGCTGCGGGAACCCACCGGGCCCTGCAG	211
QY	225	ttcgtggcatktycagattcttacgatgatggtcttaagaaagcgacgatggcaaaaac	284
Db	212	TTCTGTGGCGTTCGCGACCGCTACGACAGCGGCTGAAGCCCGCGGCGACGGGAAGCGC	271
QY	285	tacatgtcggagatgagaagcgcggaaacgaaggccttaaaagcacctctttaagctc	344
Db	272	TACATGTCCGAGATGAAACCCCGGGCTCGAGGCCCTTAAGAAACCTCTCTCTCAACAAC	331
QY	345	aacgatgtcgcgatgggaagtgaattacaatcgcggtgagctttgttggtgtactctcat	404
Db	332	GACGA-----CGTCACTTTCTCGTCTACTCCCAAC	361













Db 1101 GCGCACCCGCGGTGGCTTCTTCCTACCCAGTCGCGTTGGAATCCACGCTGGAGGCG 1160  
QY 1182 ctggcttggtggtcccggtggtgcttctcctaagtctcgatcaaggagactaatgcy 1241  
Db 1161 CTCAGCTTCGCGGTGCCATATGTTGGGATGGCGTGTGGACGAGCACGCGCACCAACGCT 1220  
QY 1242 aagatgatcaagatggtgaggaggtggtgtagagtcagagtgtaagtaggaagcggc 1301  
Db 1221 CGGACGTCGAGCTGCCTGGGCGCGGCGTGGCGCGCGCGATGCTGCGCGCGGC 1280  
QY 1302 gttgtgatagcgctgagattaaagggtgctctcgagggttataaagagtcgagag 1358  
Db 1281 GTGTTCTTCGCGGGGAGTGGAGCGGTGCGTGGCGCGCTCATGTGACGCGGCGCGAG 1337

## RESULT 9

T66173  
ID T66173 standard; cDNA to mRNA; 566 BP.  
AC T66173;  
DT 15-JUL-1997 (first entry)  
DE UDP-glucose:thiohydroximate S-glucosyltransferase clone pGL2-7.  
KW Glucosyltransferase; UDP-glucose:thiohydroximate S-glucosyltransferase;  
KW S-GT; transgenic plant; rapeseed oil; oilseed rape; canola; ds.  
OS Brassica napus cv. Westar.  
PN EP-771878-Al.  
PD 07-MAY-1997.  
PF 31-OCT-1995; 402425.  
PR 31-OCT-1995; EP-402425.  
PA (CANADA) NAT RES COUNCIL CANADA.  
PA (PLBZ) PLANT GENETIC SYSTEMS NV.  
PI Grootwassink JWD, Hemmingen SM, Kolenovsky AD, Peferoen M;  
PI Reed DW, Underhill EW, Van Audenhove K;  
DR WPI: 97-247418/23  
PT Plants genetically transformed to interfere with  
PT UDP-glucose:thiohydroximate S-glucosyltransferase gene expression  
PT - useful for production or rapeseed oil with reduced glucosinolate  
PT content  
PS Example 2; Page 18-19; 35pp; English.  
CC Brassica napus UDP-glucose:thiohydroximate S-glucosyltransferase  
CC (S-GT) partial cDNA clones pGL2-7 (T66173) and pGL2-25 (T66174)  
CC were obtd. by PCR-RACE (see also T66167-72) of B. napus cDNA.  
CC Analysis of the sequences revealed an open reading frame of about  
CC 470 bp, a 104 3' untranslated region and polyA+ tail for both  
CC clones. The amino acid sequence of the encoded protein revealed  
CC part of S-GT peptide 2 (W09827), as expected because this sequence  
CC was used for PCR-cloning, and the complete S-GT peptides 5, 6 (with  
CC one amino acid difference) and 7 (W09830-32). A full-length cDNA  
CC clone (T66166) for B. napus S-GT (W09825) was subsequently obtd.  
SQ Sequence 566 BP; 160 A; 54 C; 193 G; 159 T;

Query Match 6.0%; Score 100.4; DB 1; Length 566;  
Best Local Similarity 53.9%; Pred. No. 7.3e-16;  
Matches 234; Conservative 0; Mismatches 191; Indels 9; Gaps 1;

QY 1090 tcgtgacatggtgctcacaaatggatgttctgacgataagtcggtggatgcttgta 1149  
Db 63 TTGTTCTTGGGTAAACAGCTTGAGGTTTACGTCATGATGATCTATAGGTTGTTTGA 122  
QY 1150 cgcattgcggttggaattctgctatcgagagcctggtggtggtgcccgtggtgct 1209  
Db 123 CTCACCTGCGGTGGAACTCGACGCTGGGAAGGGTTGAGTTTGGAGTTCGCGTGG 182  
QY 1210 ttctcctaattgctatcaaggagactaatcgaaagatgatcgagagtggtgagagtg 1269  
Db 183 TGCCGCGAGTGGAGTGATCAGATGAATGATGCTTAAGTTTGTGGAGGAGGTTTGGAGAGTTG 242  
QY 1270 gttgatagtcagagtgatgaggaagcggtggtggtggtgagagtgatgagagtg 1329  
Db 243 GGTATAGACCGAAGATGAAGCTGGGGGAGAGTGTGTAAGAGCGATGAGTGGTGAGGT 302  
QY 1330 -----gcctctcgagaggttataaagatcgagagttgagataaagcgcaatgat 1380

Db 303 GTTTGAAAGAGGTGATGTAAGGAGACAGTAGTGTGAGATTAGAGAAAGTTCTTAAGAAAT 362  
QY 1381 ggaaggttttgcttaaaagactatgatgaagaaactggtgatcatcaatgaacaactctga 1440  
Db 363 GAAAGAGATTGGCTGTGAAGGCGATGAGTGAAGGAGGAAGCTCTGATCGGAGCATTAATG 422  
QY 1441 agaatttttattactagattattataaataagatgcctccataaagtgtgactatatgttat 1500  
Db 423 AGTTGTTGAGAGTTTAGGGAAGAACATGAGAGTAATGAGATTGTGAATCTTGTGT 482  
QY 1501 tattgtgttatgg 1514  
Db 483 GTTTGTTGTTGTG 496

## RESULT 10

T68693  
ID T68693 standard; cDNA; 432 BP.  
AC T68693;  
DT 05-JAN-1998 (first entry)  
DE Strawberry UDP-glucuronosyl transferase cDNA.  
KW Strawberry; UDP-glucuronosyl transferase; ERT1b; fruit; ripening;  
KW storage; gene regulation; ss.  
OS Fragaria ananassa Duch. cv. Brighton.  
PN W09721816-Al.  
PD 19-JUN-1997.  
PF 12-DEC-1996; G03076.  
PR 13-DEC-1995; GB-025459.  
PA (ZENE) ZENECA LTD.  
PI Manning K;  
DR WPI: 97-332787/30.  
PT Vectors encoding new enzymes for regulating ripening of fruit - used  
PT particularly to improve storage properties, processing  
PT characteristics etc, especially in strawberries  
PS Claim 1; Page 23-24; 43pp; English.  
CC This cDNA sequence comprises a UDP-glucuronosyl transferase (ERT  
CC 1b) cDNA clone isolated from a ripe strawberry fruit cDNA library  
CC by differential screening. 9 Ripening-enhanced clones (see T68687-  
CC 95) clones were isolated from the library, and are deposited at  
CC NCIMB 40693. Any of these 9 sequences can be used as a regulatory  
CC sequence in claimed vectors for genetic transformation of plant  
CC cells to regulate fruit ripening. The vectors also comprise a  
CC promoter and transcription termination sequence. Also claimed are  
CC plants (especially strawberry) and propagating materials containing  
CC these vectors. Stable integration of the regulatory sequences, or  
CC their complements or genomic equivalents, into a plant is used to  
CC modulate fruit ripening by overexpression or downregulation of an  
CC endogenous plant gene. Slowing ripening of fruit improves  
CC resistance to damage during harvesting, packaging and transport.  
CC It also extends shelf life, improves storage properties, processing  
CC characteristics, flavour and aroma, and modifies colour and  
CC increases resistance to post-harvest fungi.  
SQ Sequence 432 BP; 101 A; 84 C; 136 G; 94 T;

Query Match 5.0%; Score 83.8; DB 1; Length 432;  
Best Local Similarity 59.6%; Pred. No. 8.2e-12;  
Matches 136; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1083 gggaaaaatcgacatggtgctcacaatggatgtctgacgcataagtcggtggatgc 1142  
Db 112 GCGAAAGTTGTGCAATGGAGTCCACAAGAGAAAGATTTTGGAGCATCTCTCGACGGCTTGC 171  
QY 1143 ttctgacgcattgcggttggaattctgctatcgagagcctggtggtggtgcccgtg 1202  
Db 172 TTGTGACTCATTTGCGGTTGGAACTCAACCATGGAGTCACTCACTCAGGAATGCCCGTG 231  
QY 1203 gttgtcttctcctaagtgttcgataagggactaatcgaaagatgatcgaaagatggtg 1262  
Db 232 GTGGCAATTCACCAATGGGGTGACCAAGTGACCGACGCCCAAGTATTTGGTCGACGAGTTT 291  
QY 1263 aggaagtggtgtagaacatcgagtggaatgaggaagcggtgttggat 1310

V23108	ID	V23108 standard; cDNA to mRNA; 1597 bp.
AC	AC	V23108;
AD	AD	06-AUG-1998 (first entry)
DE	DE	UDP-glucose:flavonoid 3,5-O-glucosyl transferase gene.
KW	KW	UDP-glucose:flavonoid 3,5-O-glucosyl transferase; gentian; rose plant;
KW	KW	blue flower; ds.
OS	OS	Unidentified.
FT	FT	key
FT	FT	Location/Qualifiers
FT	FT	39..1460
PN	PN	/*tag= a
PN	PN	J10113184-A.

PU	23-SEP-1994.	
PF	22-MAR-1994;	G00581.
PR	23-MAR-1993;	GB-005860.
PR	22-MAR-1993;	GB-005862.
PR	22-MAR-1993;	GB-005865.
PR	22-MAR-1993;	GB-005859.
PR	22-MAR-1993;	GB-005866.
PR	22-MAR-1993;	GB-005869.
PR	22-MAR-1993;	GB-005867.
PR	22-MAR-1993;	GB-005868.

CC 5' sequences. GST is involved in the biosynthesis of steroidal  
CC glycoalkaloids in solanaceous plants, catalysing the UDP-glucose  
CC dependent conversion of the aglycone solanidine to gamma-chaconine.  
CC The invention relates to DNA sequences which encode SGT, and their  
CC use, particularly use of an antisense DNA construct to inhibit SGT  
CC activity and glycoalkaloid levels in solanaceous plants.  
CC Transgenic plants are claimed, particularly tomato and potato, that  
CC have reduced contents of toxic steroidal glycoalkaloids. The  
CC inactivation of glycoalkaloid biosynthetic pathways is beneficial  
CC to reduce or eliminate glycoalkaloid biosynthesis during shipping  
CC and storage. Fragments of SGT nucleic acids are also useful as  
CC probes for detection and isolation of SGT-encoding genes.

CC Sequence 1607 BP; 484 A; 306 C; 334 G; 483 T;  
SQ

Query Match 2.7%; Score 45.4; DB 1; Length 1607;  
Best Local Similarity 47.1%; Pred. No. 0.038;  
Matches 200; Conservative 0; Mismatches 201; Indels 24; Gaps

QY 837 aaagcggatgaactacatggaactggctgttaaactcaaaccccgcaatcatcggtggtttac 896  
|| ||||| || ||||| || ||||| || ||||| ||||| ||||| |||||  
Db 833 AACACAATGAGATTGTTATAGATTGGTGAATGCACAGAAACCTAAATCGGTTCTCTAT 892  
QY 897 gtttcgtcggagacctccttgaggctcccgcgaacccccaaatggagaataatgcaaatggg 956  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 893 GTATCTTTTCGGAAGCATGGCTAGATTTCCTCAGAGGCCAACCTCAATGAATAATAGCCCAAGCT 952  
QY 957 ctttcagaccacaattcgcaattctcttgggtgataagagaagaagagggcgacgaa 1016  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 953 CTGGATGCTTCAAATGTTCCCTTCATTTTTTGATTGAGGCCCTAATGAAGAACCAGCGCTCG 1012  
QY 1017 caagaaagcagcagaagaagaagagctgctgagctcttcttgatcgctcacggaactgaa 1076  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1013 TGGTTCCAGTGGTAATTTAGAGGCAAG-----ACTAAA 1048  
QY 1077 cgactcgggaaaaatcgtgcacatggtgctcacaaattggatgtctgacgcataagtgcggtg 1136  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1049 AAGGGTTTGATACATCAAAAGGGTGGGTCCACAGCTTAGCATCATGGAAACATTCAGCAACA 1108  
QY 1137 gaatcgttcgcacgatcggttgggaattctgctatcgagaccctggtctgttgtgtg 1196  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1109 GGCGGGTTTCATGACTCATGTGGTACTAATTCCGGTTCTGGGAAGCCATCACATTTTGGCGTG 1168  
QY 1197 ccgctggtgtcttctccaatggttcgatccaaggaggaactaatgcgaagatgacgaagat 1256  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1169 CCAATGATTAACATGGCCACTTTATGCTGATCAATTCACACGAGAAGGTAGTCGAGTT 1238  
QY 1257 gtgtg 1261  
||  
Db 1229 AGGGG 1233

RESULT 15  
V17054  
ID V17054 standard; cDNA; 1624 BP.  
AC V17054;  
DE 08-JUN-1998 (first entry)  
DT Glucosyl transferase (Grase) encoding wound inducible gene (TW11).  
DE Glucosyl transferase; Grase; TW11; tomato; signalling pathway;  
KW salicylic acid; jasmonic acid; ethylene; wound inducible gene;  
KW plant defence protein; plant response; tobacco; rice; ss.  
OS Lycopersicon sp.  
FH Key Location/Qualifiers  
FS Key 2..1410  
FT CDS /\*tag= a  
FT FT /product= "glucosyl transferase"  
FT FT /note= "encodes Grase from amino acid position 5  
FT FT shown in W47172".  
PN WO9745546-A1.  
PD 04-DEC-1997.  
PF 30-MAY-1997; G01473.  
PR 31-MAY-1996; GB-011420.  
PA (UUYO-) UNIV YORK

PI Bowles DJ, Calvert CM, ODonnell PJ, Roberts MR;  
 DR WPI; 98-032653/03.  
 DR P-PSDB; W47172.  
 PT Tomato wound inducible (TWI1) gene encoding glucosyl transferase -  
 PT useful to develop products that alter signalling pathways in plants  
 PT by altering of salicylic acid, jasmonic acid or ethylene  
 PS Claim 1; Fig 1; 52pp; English.  
 CC This wound inducible gene (TWI1) isolated from wounded tomatoes encodes  
 CC a glucosyl transferase (Grase) protein. This TWI1 gene can be used to  
 CC identify homologue Grase encoding genes isolated from tobacco and rice.  
 CC A microbial host can be transfected or transformed with a vector  
 CC containing the Grase encoding nucleic acids. The products can be used to  
 CC interfere with Grase and therefore alter signalling pathways in plants,  
 CC specifically tobacco, rice or tomato plants by altering levels of  
 CC salicylic acid, jasmonic acid or ethylene. This can induce the  
 CC production of plant defence proteins such as pathogenesis-related (PR)  
 CC and proteinase inhibitor (PIN) proteins which regulate plant development  
 CC (plant growth, reproduction and senescence) and improve plant response to  
 CC pathogens.  
 CC Sequence 1624 BP; 545 A; 252 C; 372 G; 455 T;

Query Match 2.78; Score 45.4; DB 1; Length 1624;  
 Best Local Similarity 44.6%; Pred. No. 0.039;  
 Matches 178; Conservative 0; Mismatches 221; Indels 0; Gaps 0;  
 QY 1025 agcagaagaagaagagctgctgagcttcttctgctcgcgcgaactgaacactcgg 1084  
 DB 934 AACAGGAATGAAGATTGGCTCCAGAGGATTTCAGGAAGAACACAAAGAAAGGTTT 993  
 QY 1085 gaaatcgtgacatggtgctcacatggatggttctgacgcataagtcggtggatgctt 1144  
 DB 994 AATCATAAGAGGATGGCCACCCCAAGTGTGATTCTTGATCAGCAAGCTATTGGAGCTTT 1053  
 QY 1145 cgtgacgcattgctgggaattctgtatcgagagcctggttctggtcgcctggt 1204  
 DB 1054 TGTTACTCATTTGGATGGAACTCGACACTGGAAGGAATATCAGCAGGGGTACCAATGGT 1113  
 QY 1205 gtgcttctcctcaatggttctgacgaaggactaatcgaaagatgatcgaaagatggtgag 1264  
 DB 1114 GACATGCCAGTATTTCGGGNACAGTTTTCATGAGAAAGTTGGTGTGACTGAGGTAATGAG 1173  
 QY 1265 gagtgtgtgagagtcagagtgaaatgaggaagcggcgttctgtatagggcgtgagattaa 1324  
 DB 1174 AAGTGGAGCTGCTGTGTTCTAAGCAATGGAGAGAACACAGCTAGTGAAGCAGTGCAAAAG 1233  
 QY 1325 gagggtgctcggaggttataaagagtcgagagttgagagaagcgaatgatgtgaa 1384  
 DB 1234 AGAAGCAATAGCAAAAGCGATAAAGAGAGTAAATGCCGAGTGAAGAAACAGAGGGATTTCAG 1293  
 QY 1385 ggggttgctaaagaagctatggatgaagaacgtggatc 1423  
 DB 1294 AAGCAGAGCAAAAGACTACAAAGAAATGGCAAGAGAGACC 1332

Search completed: August 1, 2000, 18:18:09  
 Job time: 20010 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 13:07:11 ; Search time 9137.45 Seconds  
(without alignments)  
326.374 Million cell updates/sec

Title: US-09-147-955-7  
Perfect score: 1671  
Sequence: 1 aacacataaaaaaaaaataa.....ctaaaaaaaaaaaaaaaaa 1671

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 972840 seqs, 892348106 residues 1945680  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 1000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: gb_bal:*	GenEmbl:*
2: gb_ba2:*	
3: gb_om:*	
4: gb_ov:*	
5: gb_pat:*	
6: gb_ph:*	
7: gb_pl1:*	
8: gb_pl2:*	
9: gb_pr1:*	
10: gb_pr2:*	
11: gb_pr3:*	
12: gb_ro:*	
13: gb_sts:*	
14: gb_sy:*	
15: gb_un:*	
16: em_fun:*	
17: em_humi:*	
18: em_hum2:*	
19: em_in:*	
20: em_om:*	
21: em_or:*	
22: em_ov:*	
23: em_pat:*	
24: em_ph:*	
25: em_pl:*	
26: em_ro:*	
27: em_sts:*	
28: em_sy:*	
29: em_un:*	
30: em_vi:*	
31: gb_htg1:*	
32: gb_htg2:*	
33: gb_inl:*	
34: gb_in2:*	
35: em_ba1:*	
36: em_ba2:*	
37: em_hum3:*	
38: em_hum4:*	
39: gb_pr4:*	
40: gb_htg3:*	
41: gb_htg4:*	
42: gb_htg5:*	
43: gb_htg6:*	

Result No.	Score	Query Match %	Length	DB	ID	Description
1	484.4	29.0	2043	7	AB013598	AB013598 Verbena h
2	446.8	26.7	1458	7	AB013597	AB013597 Perilla f
3	433.2	25.9	1507	7	AB013596	AB013596 Perilla f
4	289	17.3	1864	7	AB027455	AB027455 Petunia x
C 5	217.8	13.0	84203	49	AC005106	AC005106 Genomic s
C 6	217.8	13.0	103223	8	AC007153	AC007153 Arabidops
C 7	192.6	11.5	198944	8	ATCHRIV38	AL161538 Arabidops
C 8	192.6	11.5	200576	7	ATFCA0	297335 Arabidopsis
9	166.8	10.0	1589	8	ATU81293	U81293 Arabidopsis
C 10	163.6	9.8	195068	8	ATCHRIV42	AL161542 Arabidops
C 11	163.6	9.8	205065	7	ATFCA4	297339 Arabidopsis
C 12	162.2	9.7	197419	8	ATCHRIV41	AL161541 Arabidops
13	131.4	7.9	1594	7	AB000623	AB000623 Nicotiana
14	116	6.9	1440	49	AF190634	AF190634 Nicotiana
15	105.8	6.3	80818	7	AB019232	AB019232 Arabidops
16	103.6	6.2	568	5	A62510	A62510 Sequence 15
17	103.6	6.2	568	5	A72383	A72383 Sequence 15
18	103.6	6.2	1459	5	A62525	A62525 Sequence 30
19	102.4	6.1	94786	49	AC002333	AC002333 Arabidops
20	102	6.1	1513	5	A62523	A62523 Sequence 28
21	102	6.1	1513	5	A72396	A72396 Sequence 28
22	102	6.1	1969	5	A62529	A62529 Sequence 34
23	101.8	6.1	1416	7	M2E1AGLU	L34847 Zea mays IA
24	100.4	6.0	566	5	A62509	A62509 Sequence 14

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES





```

Qy 925 cgaaccccaatgaaataagcaatagggcttcacagacacacaaatcgcaattctct 984
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 888 CGAGTCCCAATGGAAGAGATAGCAAGAGGCTGTAGATGTGGGAGGCCGTTTTGT 947
Qy 985 ggggtgaagaagaaacgaagggcgacgacacaaagagcaagcagaagaagaagc 1044
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 948 GGGTGGTAAGAGTAACCAAGAGAGAGTA-----TTGATAAGT 988
Qy 1045 tgcagagctttcttgatcgctcacggaaactgaacgactcgggaataatcgatgggtct 1104
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 989 TGCAATGGAGGAGTTGA-----AACGAGTGGGAAAAATTGTATCTTTGGTGT 1034
Qy 1105 cacaattgatgtctcagcagcataagtcggtgggatgtcttgacgcattcggttga 1164
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1035 CTCAATTGGAAGTCTGACGCATCCCTCGTTGGGATGTTCTGACACACTGCGGGTGA 1094
Qy 1165 attctgctatcagagacctggctgtgtgtgtgcccgtgtgtgtcttctcaatggttcg 1224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1095 ATTGACTCTAGAGAGTATATCTTCGGGGTCCGATGGTGGCTTTTCCGCGAGTGTTCG 1154
Qy 1225 atcaaggactaaatcgaaagatgatcgaagatgtgtgagagagtggttgagatcagc 1284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1155 ATCAAGGACCAATGCGAAGCTGATGGAAGTGTGTGGAGACCGGCTGAGAGTGAGAG 1214
Qy 1285 tgaatgaggaaggcggtgtgtgtatagcgctgagattaaagggtgcctcggaggtta 1344
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1215 CTATGAGGAGGTAGCGCTGTTGATGTTGATGTAATTTAGGAGATGTTATTGAGGAGTTA 1274
Qy 1345 t-----aaagagtcagaggttgagaaagcgcaatgatgtggaaggtttgg 1392
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1275 TGGATGGGGAGAAAAGAGTAGGAAACTTAGAGAGAGTGTCTGGCAAGTGAAGGATTTGG 1334
Qy 1393 ctaaagaagctatgatgaagaacgttgatcatcaatgaacaaatcagaatatttatta 1452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1335 CAAGAAAAGCTATGGAGGAAGA---TGGATCTTCAAGTTAAACAACCTCAAGTCTTCTTG 1391
Qy 1453 ctaggattat 1462
    ||||| |||||
Db 1392 ATGAGTTGT 1401

RESULT 2
AB013597 1458 bp mRNA PLN 20-MAR-1999
LOCUS Perilla frutescens PF3R6 mRNA for UDP-glucose:anthocyanin
DEFINITION 5-O-glucosyltransferase homologue, complete cds.
ACCESSION AB013597
VERSION PF3R6; UDP-glucose:anthocyanin 5-O-glucosyltransferase homologue.
KEYWORDS Perilla frutescens (variety:crispa, cultivar:Shikun) leaf cDNA to
SOURCE mRNA.
ORGANISM Perilla frutescens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Gentianaceae; Lamiales; Lamiaceae; Perilla.
1 (sites)
Yamazaki,M., Gong,Z., Fukuchi-Mizutani,M., Fukui,Y., Tanaka,Y.,
Kusumi,T. and Saito,K.
Molecular cloning and biochemical characterization of a novel
anthocyanin 5-O-glucosyltransferase by mRNA differential display
for plant forms regarding anthocyanin
J. Biol. Chem. 274 (11), 7405-7411 (1999)
99167509
2 (bases 1 to 1458)
Yamazaki,M., Saito,K. and Gong,Z.
Direct Submission
Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami
Yamazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. &
Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba. 263-8522, Japan
(E-mail:mamiy@p.chiba-u.ac.jp. Tel:+81-43-290-2905,
Fax:+81-43-290-2905)
Location/Qualifiers
FEATURES

```

## source

```

1. 1458
/organism="Perilla frutescens"
/cultivar="Shikun"
/variety="crispa"
/db_xref="taxon:48386"
/tissue_type="leaf"
29. 1360
/gene="PF3R6"
29. 1360
/gene="PF3R6"
/codon_start=1
/product="UDP-glucose:anthocyanin 5-O-glucosyltransferase
homologue"
/protein_id="BAA36422.1"
/db_xref="GI:4115561"
/translation="MWRRLVLLATFPAGQHINPALQFAKRLKAGTDVTPFTSVYAWR
RMNTASAAAGNPGDLDFVAFSDGTDGLKPGGKRYMSKMGKRGSAALNLLNND
DVTFFVYSHLFAMAAEVARSHVPTALLWEPATVLCIYHFYFNGYADEIDAGNEIQ
LPRPSEQRSLPTFLPATPERFLMMKEKLELDGEAKVNLVTFDLEPDLALTA
IDRYELIGLIPSAFLDGDPSYSGDLFEKSENNKCVLNKSPKSVVVSF
GSVLRFPKAMEEIGKGLACGRPLMWIREOKNDGCEFEFEELSCIGELKMGKI
VSWCSQLEVLNLAHPALGCFVTHCGWNSAVESLSCGIPVAVPQWEDQTTNKLIEDAWG
TGVVRNMEGGVDGCEIERCVEMVMDGDKTLVRENAIKWKTIAQAMG"
1458

```

## gene

```

29. 1360
/gene="PF3R6"

```

## CDS

```

29. 1360
/gene="PF3R6"

```

## polyA\_site

```

/translation="MWRRLVLLATFPAGQHINPALQFAKRLKAGTDVTPFTSVYAWR
RMNTASAAAGNPGDLDFVAFSDGTDGLKPGGKRYMSKMGKRGSAALNLLNND
DVTFFVYSHLFAMAAEVARSHVPTALLWEPATVLCIYHFYFNGYADEIDAGNEIQ
LPRPSEQRSLPTFLPATPERFLMMKEKLELDGEAKVNLVTFDLEPDLALTA
IDRYELIGLIPSAFLDGDPSYSGDLFEKSENNKCVLNKSPKSVVVSF
GSVLRFPKAMEEIGKGLACGRPLMWIREOKNDGCEFEFEELSCIGELKMGKI
VSWCSQLEVLNLAHPALGCFVTHCGWNSAVESLSCGIPVAVPQWEDQTTNKLIEDAWG
TGVVRNMEGGVDGCEIERCVEMVMDGDKTLVRENAIKWKTIAQAMG"
1458

```

## BASE COUNT

```

353 a 348 c 438 g 319 t

```

## ORIGIN

```

Query Match 26.7%; Score 446.8; DB 7; Length 1458;
Best Local Similarity 63.0%; Pred. No. 9.3e-99;
Matches 896; Conservative 2; Mismatches 434; Indels 90; Gaps 10;
Qy 57 cgcataatctactagcaaatcccccagcaagaagccacataaaacccctctctcagtttc 116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38 CGCCCGCTGCTCTAGCAACGTTTCCGGCGCAAGGCCACATAAATCCCGCCCTCCAATTC 97
Qy 117 gccaaaagctctcaacacacgggatacgtgcaccaaagtcacattcttcacagtgatatac 176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98 GCCAAGACACTCTCTAAAAGCCGG-----CACTGAGTCACGCTTTTTCACGAGCGTTAT 151
Qy 177 gcattgagacgcgcgtctcgaaacccgatccgagcagcagaa-----tcgat 224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 GCATGGCGCCGATGCCCAACACACAGCCTCCGCGCTGCGGAAACCCACCGGCGCTCGAC 211
Qy 225 ttcgtggcatktycagattcttacgatgatgcttaagaagaagcgagcgatggcgaataac 284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 TTCGTGGCGTTCTCCGACGGCTTACGACGACGCGGCTGAAGCCGCGGCGGACGGAAGCGC 271
Qy 285 tacatgctggagatgaaaaacgcggaagcgttaaaaggacacattcttataagctc 344
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 272 TACATGTCGAGATGAAGCCGCGGCTCCGAGGCGCTTAAGAAACCTCTCTCTCAACAAC 331
Qy 345 aacgatgctggatgggaagtgaatgttacaatcgctgagcttctgtgtgtactctcat 404
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 332 GAGCA-----CGTCACCTTCTGCTGCTACTCCCAC 361
Qy 405 ctatttctgtggcgagctgaagtggcggtgaagtgcagtcgagtcgagtcgcttcttgg 464
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 362 CTCCTTTGTATGGCGCGGAGGTGGCGGCTTTGTCCACAGTCCCGACCGCCCTCTCTGG 421
Qy 465 attgaacggctacggttttcttgatgttactatttttctcaatgggtatcgatgat 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 422 GTCGAGCGCCGACCGTCTGTGCATATACCACTTCTACTTCAACGGCTTACGACGACGAG 481
Qy 525 atcgatcgggctcagatcaaatcccaactcttcgcagctctcccagcaagat 584
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 482 ATCGAGCGCGGTTCGAATGAATTCAGCTCCCTCGGCTTCCATCCCTGGAGCAGCGCACT 541
Qy 585 ctcccccttttctactactctcgagcccccgagatccgaacccctaatgaagaagaaag 644
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 542 CTCCTCGAGCTTCTGCTGCTCGACGCGGAGAGATTCGGGTGATGATGAAGGAGAAG 601

```

```
QY 645 ttgcacgcctgcacaaagcgaagcgaagctgttataaaccagcttcgacgcat 704
Db 602 CTGGAACTTTAGACGCTGAAGAGAGCGAAGATATTGGTGAACACTTTGATGCGTTC 661
QY 705 gaacgcgaacaaactcaagccatcgacaggtgatgaactaatatccatcgcccatatc 764
Db 662 GAGCCCGATGCACCTACGCGTATTGATAGTATGAGTTGATCGGATCGGCGCTTGATT 721
QY 765 ccatacgcgatattctcagatggcgaacaccctcatcaagaacaaactctactgtga 824
Db 722 CCTCCGCG---CTTCTGGACGCGGAAGATCCCTC-----CGAAACGCTTTACGCGCGC 772
QY 825 gacctcttcagaaagccgatgaact---tacatgagctgggttaactcaaacaccgaa 881
Db 773 GATCTTTTCGAAAAATCGGAGGAGATAACTGCTGGAGTGGTTGACTCGAAGCCGAAA 832
QY 882 tcactgttcttacgttttcgttcggagcctctcaggtctcctcggtgataagaagaac 1001
Db 833 TCCTCGCTGCTGTATGCTGCTTGGAGCGCTTTTGAGGTTTCCAAAGGCACAAATGGAA 892
QY 942 gaaatagaatagggcttcagacacacaaatcgccaggtctctcggtgataagaagaac 1001
Db 893 GAGATTGGAAAGGGCTATTAGCTCGGAAGCCCTTTTATGATGATACGAGAACAG 952
QY 1002 gaagagggcgacgaacaaagacagacagagagagagagagctgctgagctcttctgat 1061
Db 953 AAGAATGACGACGCGGCAAGAGAGAGAGAGAGAGAG-----TTGAGTTGC 1000
QY 1062 cgtcaggaactgaagactcggggaacacgtgacatggtgctcacaattggatgtctg 1121
Db 1001 ATTGGGGAATTGAAAAAATGGGGAATAATAGTCTGCTGCTGCTCGCAGTTGGAGGTTCTG 1060
QY 1122 acgcataagtcggtgagctcgtcgacgacgattcggttggaactctcgtatcgagagc 1181
Db 1061 GCGCACCCCTCGTGGGATGTTTCGTGACGCAATGTGGGTGGAACTCGGCTGTGGAGAGC 1120
QY 1182 ctgctctgtgtgctgcggtggtgctcttcccaatggttcgacaaagggactaatg 1241
Db 1121 TTGAGTTGGGGATTCCGGTGTGGCGGTGCCGACGTGTTGATCAGACGACATGCG 1180
QY 1242 aagatgcgaagatgtgtgagagagtggtgtgagagtcagagtgatgaatgaggaagcgc 1301
Db 1181 AAGCTATTGAGGATCGTGGGACAGAGGGGTGAGAGTGAGATGAATGAAT---GAAGGGGT 1237
QY 1302 gttgtgataaggtcagatgaagaggtcgctcagaggttat-----aaag 1349
Db 1238 GGGTGTGATGATGTGAGATAGAAAGGTGTGGAGATGTTGATGATGGGGGTGACAG 1297
QY 1350 agtcgagagttgagaaagcgaatgatgtggaaggggttggctaaagaagctatggtat 1409
Db 1298 ACCAACTAGTGAGAGAAAATGCCATCAATGGAAGACATTGGCCAGACAGAGCCATGGGA 1357
QY 1410 gaagaacgtggtatcattgaacaaatctgaagaattttatt 1451
Db 1358 TAGGA---TGGATCTTCACTCAACAATCTCAAGCGCTTCTT 1396

RESULT 3
AB013596 1507 bp mRNA PLN 20-MAR-1999
LOCUS Perilla frutescens PF3R4 mRNA for UDP-glucose:anthocynsin
DEFINITION 5-O-glucosyltransferase, complete cds.
ACCESSION AB013596
VERSION AB013596.1 GI:4115558
KEYWORDS UDP-glucose:anthocynsin 5-O-glucosyltransferase.
SOURCE Perilla frutescens (variety:crispa, cultivar:Shikun) leaf cDNA to
        mRNA.
ORGANISM Perilla frutescens
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
        Asterales; Gentianales; Lamiales; Lamiaceae; Perilla.
REFERENCE 1 (sites)
AUTHORS Yamazaki,M., Gong,Z., Fukuchi-Mizutani,M., Fukui,Y., Tanaka,Y.,
```

```

Kusumi,T. and Saito,K.
Molecular cloning and biochemical characterization of a novel
anthocyanin 5-O-glucosyltransferase by mRNA differential display
for plant forms regarding anthocyanin
J. Biol. Chem. 274 (11), 7405-7411 (1999)
99167509
2 (bases 1 to 1507)
Yamazaki,M., Saito,K. and Gong,Z.
Direct Submission
Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami
Yamazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. &
Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan
(E-mail:maniy@p.chiba-u.ac.jp, Tel:+81-43-290-2905,
Fax:+81-43-290-2905)
FEATURES             Location/Qualifiers
     source            1..1507
                     /organism="Perilla frutescens"
                     /cultivar="Shikun"
                     /variety="crispa"
                     /db_xref="taxon:48386"
                     /tissue_type="leaf"
     gene              17..1399
                     /gene="PF3R4"
     CDS                17..1399
                     /gene="PF3R4"
                     /codon_start=1
                     /product="UDP-glucose:anthocynsin 5-O-glucosyltransferase"
                     /protein_id="BAA36421.1"
                     /db_xref="GI:4115559"
                     /translation="MVRRLVLLATFPAGHINPQLQFAKRLKAGTDVFTFSVAVR
                     RNTASAAAGNPGLDFVAFSDGDLKPCGDKRYMSEMKARGSAIRLLNNH
                     DVFYVYSHLFAWAEARESOVSALLWEPATVLCIYFYFNGYADIDAGDEIQ
                     LPRPLQORSUPTFLPETPFRFLMKKLETLDEBEKAKVLVNTFDALPDLALTA
                     IDRYELIGLIPLSAFLDGDPSETSYGDLFEKSENNCENWLDLPKSSVTVSF
                     WGSQLEVLHAPALGCFVTHGWNVAVESLSCGPVAVVQWPDQTTNAKLIEDAWGTG
                     VVRNMEGGVGDSBIERCVENMVDGGEKSLVRENAIKWTKLAREAMGEDGSSLKNL
                     NAFLHOVARA"
BASE COUNT      366 a   351 c   448 g   342 t
ORIGIN
Query Match      25.9%; Score 433.2; DB 7; Length 1507;
Best Local Similarity 62.5%; Pred. No. 2e-95;
Matches 889; Conservative 2; Mismatches 435; Indels 96; Gaps 10;
QY 57 cgccattcttactagcaacattccacagacacagggccacataaacctctctcagattc 116
Db 26 CGCGCGTGTCTAGCAACGTTTCTCGCAAGGCACATAAATCCGCGCTCCAATTC 85
QY 117 gccaaaggctctcacaacaccggatagctgcacaaagtcacattcttccagagtgatac 176
Db 86 GCCAAGAGACTCTCTAAAGCCGG-----CACTGACGTACATTTTTCAGGAGGTTTAT 139
QY 177 gcaatgagcgcgcatcgcttcgaaacgcatcccgagcagcagaa-----tcgat 224
Db 140 GCATGCGCGCATGGCCCAACACAGCGCTCGCGCGCTGCCGGAACCCACCGGCGCTCGAC 199
QY 225 ttctgtgcatktycagattcttactgatgctttaaagaagggcgacgatggcaaaac 284
Db 200 TTCTGTGGCGTTCTCCGACGGCTACGACGAGGGGTGAAGCCCTGCGCGGACGGGAAGCGC 259
QY 285 tacatgcggagatgagaaagcgcggaagagcctttaaagagacactcttataagctc 344
Db 260 TACATGCCAGATGAAGACCCCGCGCTCCGAGGCCCTTAAGAAACCTCTTCTCAACAAC 319
QY 345 aacgatgctgcgagtggaagtgatgttacaacatcgctgagcttgtgtgtactctcat 404
Db 320 CACGA-----CGTCAGTTCGTCGCTACTCTCCAC 349
QY 405 ctatttcgtggcgagctgaagtggcgctgaagtcgacgctgcgagtcgacctctttg 464
Db 350 CTCTTTGCATGGCGGGAGGTGGCGGTGAGTCCAGGTCCGAGGCGGCGCTCTCTG 409
```

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITILE  
JOURNAL

Qy 465 attgaacgcgtacggttttcgtatgtactatttttactcaatgggtatgccgatgat 524  
Db 410 GTCAGACCGCCACCGTCTGCTGCATATATATTACTTCTCAACGGGTACGACACGAG 469  
Qy 525 atcgatcgggctcagatcaaatcaactgccaatcttcgcagctctccaagaagat 584  
Db 470 ATCGACCGCGTTCGAGCAAAATTCAGTCTCCTCGCTTCCACCCCTGGAGCAGCGCAT 529  
Qy 585 ctccctcttctactccttcgagcccgagatccgaacccctaatgaagaagaag 644  
Db 530 CTTCCGACCTTCTGCTCCGAGACACCGGAGAGATTCCGGTGTGATGATGAAGGAGAG 589  
Qy 645 ttgcacacgctcgcacaagaacccgaagcgtcttgataaaacacgttcgacgcat 704  
Db 590 CTGGAACCTTAGACGGTGAGAGAGCGGAAAGTGTGTTGGTGAACACGTTTGATCGGTG 649  
Qy 705 gaacccgaacactcaaacgcatcagcaggttatgaactaatatccatcgcccatatc 764  
Db 650 GAGCCGATGCACCTACGCTATTGATAGGTATGATGTTGATCGGATCGGCGGTGATT 709  
Qy 765 caatcatcgatattctcagatggcaacgacccctcatcaagcaacaataatcctacgttga 824  
Db 710 CCTCCGC---CTTCTTGACGGCGGAGATCCCTC-----CGAAACGTCTTACGGCGC 760  
Qy 825 gacctcttcagaaagccgatgaactta---catggactggctaaactcaaacccgaa 881  
Db 761 GATCTTTTCGAAATTCGAGAGAGATTAATCGCTGGTGGAGTGTGGACGACGAGCGAA 820  
Qy 882 tcatcggtcgtttacgtttcgttcgggagcctcctgaggtcccgaaaccccaaatgaa 941  
Db 821 TCTTCGGTGTATGTGCTGTTGGAGCGTTTTGAGGTTTCCAAAGCAGCAATGGA 880  
Qy 942 gaaatagcaataggcttttcagaccccaaatccagttctcttgggtgataagaagaac 1001  
Db 881 GAGATTGGGAAGGCGTATTAGCCTGCGGAGGCGGCTTTTATGATGATACGAGAAGCAG 940  
Qy 1002 gaagagcgacgaacgaagcaagcagaagaagaagaagagctgctgagctctttgat 1061  
Db 941 AGAATGAGCAGCGGAGAGAGCAAGCAAGCA-----GTTGAGTTGC 982  
Qy 1062 cgtcaggaactgaacgactcgggaaaaatcgtgacatggtgctcaaatgtgattcttg 1121  
Db 983 ATTGGGAATTCAAAAAATGGGAAATAGTTTCGTGGTCTGCAGTTGGAGCTCTG 1042  
Qy 1122 acgataagtcggtgggatgcttcgacgcaattcggttggaattctgctatcgagac 1181  
Db 1043 CGCACCTCGCTTGGGATGTTCTGTCGACCATTTGCGGTGGAACTCGGCTGTGGAGAGC 1102  
Qy 1182 ctggctgtggtggtccggtggtgcttctcctcaatggttcgatacgaaggactaatcg 1241  
Db 1103 TTGAGTTCCGGGTTCCGGTGTGGCGGTGCGCAGTGTGTTGATCAGCAGCAATGCG 1162  
Qy 1242 aagatgacgaagatggtgaggaggtggtgagagtcagagtgaatgaggaagcgcc 1301  
Db 1163 AAGCTGATTGAGGATGCGTGGGGACAGGGGTGAGAGTGAGATGAAT---GAAGGGGT 1219  
Qy 1302 gttgttgataggcgtgagattaaagggtgcgtctcggagggttat-----aaag 1349  
Db 1220 GGGGTTGATGATCCTGAGATGAGAGAGTGTGTGGAGATGGTGTGATGGTGGGAG 1279  
Qy 1350 agtcgagagtggagagaagcgaatgatgtggaagggttllggctaaagaagcctatgat 1409  
Db 1280 AGCAACATAGTGAGAGAAATGCCATAAATGGAAAGACTTTGGCCAGAGAAGCCATGGA 1339  
Qy 1410 gaagaacgtggtatcatcaatgaacactcgaagaattttatt 1451  
Db 1340 GAGGA---TGGATCTTCACTCAAGAAATCTCAACGCGCTTCTT 1378

RESULT 4  
AB027455  
LOCUS

07-JAN-2000

PLN

mrna

1864 bp

AB027455

DEFINITION Petunia x hybrida PH1 mRNA for anthocyanin 5-O-glucosyltransferase, complete cds.  
ACCESSION AB027455  
VERSION AB027455.1 GI:5683051  
KEYWORDS anthocyanin 5-O-glucosyltransferase.  
SOURCE Petunia x hybrida (cultivar:Surfinia) corolla cDNA to mRNA.  
ORGANISM Petunia x hybrida  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Petunia.  
1 (bases 1 to 1864)  
Yamazaki,M. and Saito,K.  
TITLE Anthocyanin 5-O-glucosyltransferase  
JOURNAL Published Only in Database (2000) In press  
REFERENCE 2 (bases 1 to 1864)  
Yamazaki,M. and Saito,K.  
TITLE Direct Submission  
JOURNAL  
SUBMITTED (17-MAY-1999) to the DDBJ/EMBL/GenBank databases. Man1  
Yamazaki, Chiba University, Faculty of Pharmaceutical Sciences;  
Inage-ku Yayoi-cho 1-33, Chiba, Chiba 263-8522, Japan  
(E-mail:mamiy@p.chiba-u.ac.jp, Tel:81-43-290-2905,  
Fax:81-43-290-2905)  
FEATURES  
source  
1. 1864  
/organism="Petunia x hybrida"  
/cultivar="Surfinia"  
/db\_xref="taxon:4102"  
/tissue\_type="corolla"  
100..1506  
/gene="PH1"  
100..1506  
/gene="PH1"  
/codon\_start=1  
/product="anthocyanin 5-O-glucosyltransferase"  
/protein\_id="BA089009.1"  
/db\_xref="GI:5683052"  
/translation="MVQPHVLTTPAQGHINPALQAKNLVKMGIEVTFSTSIYAQS  
RMDKSLILNPKGLNFIPEFSGDEGPDHSDPVFMSOLRKCSETVKKLIILTCSEN  
GQPTICLLYSIFLPWAAEVAREVHIPALLWSQATILDIYFNFHGYEKAMNESND  
PNWSITLPLLETRDLPSELPYGAAGSLVAPLPKELIDITLDAETTPKILVNTF  
DELPEALNAIEGYKFGYGLIPSAFGLGNDPLDASFGDLFONSDNYMGLSKPN  
SSVYVIFSGILMNPISQMEBISKGLIDIGRPFVWIKENKGEENKKGICIELE  
KIGKIVPCSOLEVLKHPSLGCFVSHCGWSALSACGVPVAFPOWTDQNTAKOV  
EDVWKSQVVRVINGEDVVESEIRKCIELVMDGKGKELRKNKKWELAREAVKES  
GSSHNKLFADIDYAKGF"  
BASE COUNT 603 a 318 c 369 g 574 t  
ORIGIN  
Query Match 17.3%; Score 289; DB 7; Length 1864;  
Best Local Similarity 56.1%; Pred. No. 3.5e-60;  
Matches 847; Conservative 2; Mismatches 557; Indels 105; Gaps 12;  
Qy 36 aaaaaaaaaatggttaacaaacgcgcattctactagcaacattcccgacacaaagccac 95  
Db 88 AGCAGAGAAATATGGTCAGCCTCATGTCTATCTTAACAACATTTCCACGACAGGCCAT 147  
Qy 96 ataaaccttctctcgtagttcgccaaaaggctcctcaaacccggtatacgtcgaccaagtc 155  
Db 148 ATTAATCAGCAGCTCAATTTGCCAAGAATCTTGTCAAGATGGGCATAG-----AAGTG 201  
Qy 156 acattcttcacgagtgatgatacgcattgagacgcgcgttcctogaacgcgataccgagc 212  
Db 202 ACATTTTCTACAAGCATTTTATGCCCAAGCCGTATGGATGAAAAATCCATTTTAAATGCA 261  
Qy 213 ---agcaaatcgtattcgtggcatktycagattcttacgattcagtgatggcttaagaagc 269  
Db 262 CCAAAGGATTGAATTCATTCATTCATTCCTCCGATGGCTTTGATGAAGGTTTGTATCATCA 321  
Qy 270 gacgatggcaaaactacatgctcgagatgagaaagcgcggaacgaagccttaagagac 329  
Db 322 AAAGACCCCTGATTTTACATGTCACAACCTTCGTAATGTGGAAGTGAAGTGTCAAAAAA 381



```

T7 end of YUP8H12"
/codon_start=1
/evidence-not_experimental
/product="T25N20.1"
/protein_id="AAC80580.1"
/db_xref="GI:3935137"
/translaton="MGKKEQDHSVVESDDKVEAVLHLLKHSPLTLKQEKFCNRAC
VGFRLRIGDNVAKAQLRSCLSRWSSGLIESLIADEFTAELAEGLAYVAGLDDECR
PVLVERKODYQKLTOKOLTRLVVFTLEVAISTMSRNVEQFVLDPDAIAAVALVTL
IVLESPLNLLCNSTAHLPFFFFFFFYKYNPKLLOJISFHEHTCDYTONSRBG
IRFVDSLTAIVSMQDFQDFDYSFSSYRSVSLRSLTSSLSKDTKIGSCASS
RFATVRSGLDTPKMCWLTDTSTSKLGHNTGAYISPLNARSFSPASPAKSEPPG
GPRSFATPMPTDRHSIGTLRDPRIPEPFFQSPAIFFRRSHSVKSKPRDS
FVQFLKFRPYDEMYSKMRPPLGLGVIVSQIRRHVLSLSQR"
join(9304, .11268, .11387, .11513, .11607, .11692, .12103, .12341,
12698, .12784, .13257, .13356, .13451, .14263)
/note="unknown; similar to ESTs dbj|C26511 and
gb|AI054630"
/codon_start=1
/evidence-not_experimental
/product="T25N20.2"
/protein_id="AAC80581.1"
/db_xref="GI:3935138"
/translaton="MEGGGSGERSRIASNTPSLKEKSDGVGRTVSSTETRONYQR
SRVVSDESDEFMPPRRSGVDRTLGAKKEFKRKDRVDEHRNGYVRRNNEASG
SFMVKNLDTFEFDEYDGFDSANLKRKFDNGSVGRSSFSASRRDVSQVGRSGSR
EGLFDRRRNTFVNGTCSASSQSDSDSEPMRVQGINGLVKVYNKNTNTLAASI
NPDAEIYRPPSRKAQRNVVYPPRPSKNVDNNSSESMSRKRKRKSEYS
KPKREFNTKSTPELVNPDVREERKRGGDGDKORLERIKMLTDAGWTIDYKP
RRNQSLDAYVNPSTPAWYSIIKAYDALLQKDEVDARPKDAAVASVEEIVN
KLARKAKTSEMTKKWQNSGSDSENKSEGGAYTDTSEERIRSSIKLGCKTKKGR
NGADMBLHKRSRSLYNNARPSCGSDSHYLGRTKTKIGRCTLLVRSCKKKNPAI
NPNFYPSKRTLLSWLTESGVOLROKQVYMRRGAKVLEGWITREGJHCDCCKTL
TVSEFELHAGSKSCOPONTLYLESGASLLQCYRANWMDATNLAHQVDTGGDDN
DDAGICGCGDGLCCDCSTTHQNCIGMQVLPSGDWHPNCTKCDAAVASGND
GNFTLSLSCGERRYHQLCNDDEAHKVGSGSASFCPKLELEFKLQGLVTE
IEGYSWSLIRHVDTSQNSQAQRIENNSKLAIVLAIMECEFLPIVDRRSGDLI
RNVLYNGDSNRINTFYTALIERGEDEIIISAASFHGMQLAEMPFICTRHIYRQ
GMCRLFPDAIESAMRSUKVKELVIPAIDFLHAWTGNFTPLDDSVRKEMRSINTLV
PFGIDMLQKPLHHEENIIAAGADAMISEVEEKKSEPTSSVEIGPYAVEGDEFVAD
AANCYKIDLASDENILVSVETAGTICKPKDELRSFRGEESGSSSPCCQITLKSGT
KVLHGHTDGTGSCEDGLTDVNEADSLLSQETQOASAFKVENNLISLSIGRGS
DLSSISOEVSEQTSNLDGVPCKDYNILPQAKLDKSKDDAFADGFLI"
complement(join(14820, .14957, .15038, .15420, .15643, .15721))
/note="hypothetical protein"
/codon_start=1
/evidence-not_experimental
/product="T25N20.3"
/protein_id="AAC80582.1"
/db_xref="GI:3935139"
/translaton="MGLVAVMFSPTLVHHMKSFKQCONELKRGQTSLSLDRRGF
LKCWVGASSPMATIEFSGLAQASEEKLDGEGVGAFTLFDPNERTKSGKELPRAY
LKSAREVYVYMRSLKNPKDNKAFKSDAAKESIRYLSNWRGQKTVAEGEYSVEL
ENVIRALAKYSKAGSPAPLPDEVKTEILDOLNKAEEFL"
join(16409, .16661, .16979, .17133, .17234, .17258, .17346, .17362)
/note="unknown; similar to EST emb|Z26799"
/codon_start=1
/evidence-not_experimental
/product="T25N20.4"
/protein_id="AAC80583.1"
/db_xref="GI:3935140"
/translaton="MMDPPPTKEFALAPVYVYWMKRCVPPDDYDARRVGPCKRIL
RKSGYNGPVITAVGSLSKVPRDILEVVSSTGISLYHEVAPRLEERDESSETPGPAS
WICSVRCRIGPGIAGGVDFNTHLSRHEUKLRQRTNPKGHLAN"
complement(join(17577, .18026, .18231, .18796, .18981, .19410))
/note="similar to RNA polymerase II elongation factor
homolog (S59773)"
/codon_start=1
/evidence-not_experimental
/product="T25N20.5"
/protein_id="AAC80584.1"
/db_xref="GI:3935141"
/translaton="MFLHVVLIEAASIGFIECRSYNMSDMDIQSNIGESPHLLLR

```

```

VSPLESGEGLPYAPENPNPGDTWHWKVGPRIKSGYFVDVRYLYPPKYLPLGLDTEILR
KNKVFSSRLSLQRYIRVHEADYQKEFFASFSWSIPLCRDGGVLPQKQVOLPVYSSDE
DPMRDGSDTAVCKAGNEKRSCLMPQCEATLPMFCDICCGBERKFCYDCCLLCCKL
ISLEHGYSTIKCEAVYSEGHICGHVAAEMTCLRAYLAGTIGSGMDGTETCYRRCDA
KDLDFHVNKFLCIEQTVYOGDVETILNLGICILRGAQRNAKELLNCLNCEJESTYKII
FIIYVDKSFSEARENDTQSLQDVTPIGPIFNHEAEMHLEEEIGEVULARKKAQE
FEYQIABGKQAGCECLSDYROLEKEKSELSRRVSGTDSANSLMTNLVKRLDQIRKEV
TKLKEVEAKFGCGRTPRGVLEEYFHLNIED"
22998, .22826
/note="unknown; similar to ESTs emb|Z33701 and
emb|Z34709"
/codon_start=1
/evidence-not_experimental
/product="T25N20.6"
/protein_id="AAC80585.1"
/db_xref="GI:3935142"
/translaton="MLNTYSLTLFTLTKTMRVMMKNFHLCPFSNLTKSPSSPSGAT
SDDPNRPSILLNNFLNYDDSSAAHRLSKPLIHDEPSTFASTSTAANSSSA
SYDDSDNYGAPDDSPDPDLTAVLASRRFFSPGCGSNSITSDPDLCRCDNYATNR
LLTGSTAVKHVQSPDPYNDPRRSMQEMIDAVTNAGDLRVRVEFLHELHLLSYLSNAAD
THKFIIRAFADILVSLSDGHRIS"
complement(join(26300, .26503, .26613, .26880, .27004, .27259,
27394, .27499, .27748, .27804))
/note="unknown; similar to ESTs gb|T76156 and gb|N96012"
/codon_start=1
/evidence-not_experimental
/product="T25N20.7"
/protein_id="AAC80586.1"
/db_xref="GI:3935143"
/translaton="MAMALNOLFVVLASKPEQKITPEESRAIVSCHFALMTAGFA
SGVGLLQWTKTKLPGKGLERVALAGNYLHIFVYDHSFRLSEKCECVFFSFTSQY
ELLRLALLIGIGVQPNMOPHLIISLVKMLHECRKFLVRSNREGRAWVFFSKHFY
PESVYQDEGDKPOMWRERRTFEIASYDDVNATKSQPNPGI.PNPSHRIISGSDA
SKYQTLQNSGSDGEMAEEDVLIDVFGSEATESIPAPVISLASKTQTRKQRAQ
RRRLKKNREASTNTPQYELA"
complement(join(28447, .28740, .28976, .29125, .29206, .29943))
/note="hypothetical protein"
/codon_start=1
/evidence-not_experimental
/product="T25N20.8"
/protein_id="AAC80587.1"
/db_xref="GI:3935144"
/translaton="MKSARANOMINRETMKGKILVSPSEKSMSTSTTTTTTTTTVS
TEWERRDSNCPYPPDKRDANDSCIECLDSLNATLDMPLSVQKSLTKLSASNFKS
TVESTPSPDPTVYVTPASVSRPILKLMISPPKKLTKKSKVFENEQRTKERRRL
LIVVFLAVIIVIGLVGLGELGSVWLEVEFPEFTEIVRNAYEQADLDLGVKRL
LEDEFKGFVTSRKPSRCTSDSKWKINODGPMKNSKCVLYKSAIEEYSIWGPLOTSG
LFTFTGSSSITVLSGRVTEGKFGYSIHTNTSWGKTWSTVLQDPTWVLEY
SQSSVMDSSLSLLSLTIDIMEHLVFRVAKVNRERYWMPSSSRRLYEAESKASATP
T"
join(32077, .32407, .32715, .32956)
/note="similar to lipid transfer protein (AB007843)"
/codon_start=1
/evidence-not_experimental
/product="T25N20.9"
/protein_id="AAC80588.1"
/db_xref="GI:3935145"
/translaton="MNSNELISAALIFSLSSNSPTSILAQINPCSPSMLSVTGC
TSFLTGGGSFPTSDCCAKSLTGTGMDCLCLIVTAGVPISIPINRLAISLPACGI
PGVPVCKPGFPTTSPDTSQTSDEGTRELEFATSSSSSKTFCSEFVSKRFSGLL
TSNHSQGLLSVRPRLRLKRLKLMVNHNS"
complement(33000, .33155)
/note="similar to EST emb|Z26808"
/codon_start=1
/evidence-not_experimental
/product="T25N20.10"
/protein_id="AAC80589.1"
/db_xref="GI:3935146"

```

Query Match 13.0%; Score 217.8; DB 49; Length 84203;  
 Best Local Similarity 54.2%; Pred. No. 1,1e-42;  
 Matches 775; Conservative 2; Mismatches 564; Indels 90; Gaps 13;

Qy	35	aaaaaanaaaatggtttaaaaaacgccatattcttactagcaacattccacgacaaagccca	94
Db	79169	AAGTAGAAAAAATATGCGCCACGCAATTTTCTACTGTAACTGTTCCGGCGCAAGGTCA	79110
Qy	95	cataaaccttctcgcagttcgcgaagaagcctcctcaacacggatagctcgaccgaagt	154
Db	79109	CGTGAACCATCTCTCGGTTTGTCTGCTGCGCTCATCAAAA---GACCGGCGCAGCTGT	79053
Qy	155	cacattcttcacgagtgatacgcgcatigagcgcgcatgccttcgaaaccgcatccgagcag	214
Db	79052	CACTTTGGCTACTTGTGTCCTCGTCTCCCAACTCCATGATCGCAAAACCAACAAGT	78993
Qy	215	cagaa---tcgattctggtgcatkyacagattcttcagatgatggcttaaagaagcgga	271
Db	78992	CGAAATATCTCTCTTCTTACTTTCCTCGACGGTTTCGACGATGGAGCATTTCCACCTA	78933
Qy	272	cgalggcaaaaactacatgctcgagatgagaaagcggaagcggaagcctttaaaggcac	331
Db	78932	CGAAGACGTCAGAAAAGTGGTGAATCTCAA-----GGTTAACGGCGATTAAGGCACTA	78878
Qy	332	tcctataagctcaacgatgctcgatgggaagtgaatttcacaatcgctgagctttgt	391
Db	78877	TCCGATT---TCATCGAAGCTACTAAGATGGTGACTC-----TCCCGTGACTTGCTT	78828
Qy	392	gggtactctcatctattttcgtgggcagctgaagtggcgctgaagtcgacgtgcgcgag	451
Db	78827	GATCTACAGATTCTTCTCAATTGGGCTCCAAAAGTAGACACGTAGATTTCAACTTCCTC	78768
Qy	452	tgcctctcttgagattgaacccgctacggcttttcgatgtgtactattttacttcaatgg	511
Db	78767	CGCTTCTCTCGSATCCAAACCGGCTTGGTTTTCAACATCTATTACACTCATTTTCATGGG	78708
Qy	512	gtatgcgcgatgatcatcgatcgggctcgatcctcaaatccaaactgcctcctccgcagct	571
Db	78707	AAACAAGTCCGTTTTCCAGT-----TACTTAATCTGTCTTCTCT	78659
Qy	572	ctccaagcaagatctcccctcttctactcccttcgagcccgcgagattccgaaac--	629
Db	78668	GGAAATCAGAGATCTTCCATCTTCTCCACACCTTCCAAACAAACAAAGCGCATACGA	78609
Qy	630	-ctaattgaagaagaattcgacacgctcgacaaagaacccggaagcgaggtcttgataaa	688
Db	78608	TGCGTTTCAAGAAATGATGGAGTTTCTCATAAAAAGAAACCAAAACCGAATAATTCATCAA	78549
Qy	689	cacgtttcgacgcatatgaacaacgcaacaactcaaaagccatcgacaggtatgaactaatc	748
Db	78548	CACTTTCGATTCGCTGGAAACAGAGGCCCTTAACGGCTTCCCGAATATCGATATGGTGGC	78489
Qy	749	catcgcccaattaatcccatcatgatattctcagatggcaacgaccctcatcaagcaa	808
Db	78488	GGTTGGTCTTACTTCCACGGAGATTTTCTCAGGAAGCAACCAAAATCA-----	78435
Qy	809	caaatcctacggttgagaccctctcagaanaagccgatgaaaacttacatggaactggctaaa	868
Db	78436	-----GTTAAAGATCAAAAGTAGTAGTTATACATTTTCGCTAGA	78399
Qy	869	ctcaaaaccggaaatcatcgtgctgttacggtttcgttcggagacctctgaggtcccgaa	928
Db	78398	CTCGAAACACAGTCCCTGTATTATTACGTTTCTTGGAAACANTGGTTGAGTTGTCCAA	78339
Qy	929	acccaaatggaagaaatagcaatagggtcttcagacaccaaatcgccagttctctgggt	988
Db	78338	GAACAGATAGGAACTAGCGAGAGCACATATAGAAGGGAACACCGTTTTGTGGGT	78279
Qy	989	gataa---gaagaacgaagagggcgacgaaca-agaagcaagcagaagaagagaagct	1045
Db	78278	TATAACTGATAAATCCAAACAGAGAAACCAAAACAGAGAGGAGAAAGACAGAGATGA	78219
Qy	1046	gctgagctctttgatcgtcacgcaactgaacgactcggaagaatcgtgacatggtctc	1105
Db	78218	GAAGATAGCTGGATTACACACAGAGCTTGAAGAGTTGGAGTTGTGTGCTGGTGTTC	78159
Qy	1106	acaattggatgttctcgacgataaagtcgggtgggagtgctcgtgacgcatctcgattgaa	1165

Db	78158	GCAGATAGAGTTTAAAGTCAACGACCGTAGTGTGTTTGTGACTATTGCGGTGGAG	78099
Qy	1166	tctgctatcagagcctgcttgggtgtcccgtagtggcttctctcctcaatggttcga	1235
Db	78098	CTCAGCGCTGAGAGATTGTTGGTCTTGGCGTTCCGGTGTGCGGTTTCGGATGTGGTCGGA	78039
Qy	1226	tcaaggagactaatcgcaagatgatcgaaatggtgagggagtggtgtgagagtcagagt	1285
Db	78038	TCAACCGACGAACCGGAAGCTACTTGAAGAAGTTTGAAGAACTGGTGTGAGGGTAAGAGA	77979
Qy	1286	gaatgagaaagcggcgcgttggtagatagcgtgagattaagagggcgctctcgcggggttat	1345
Db	77978	GAACAAGGA---TGGTTTGGTGGAGAGAGAGATCAGGAGGTTGTTTGGAAAGCGCGTGT	77922
Qy	1346	-----aaagagtcgagagttgagaaagcgaatgatgtsgaaggggttggcctaaaga	1399
Db	77921	GGAGAGAGATCGGTGGAGTTTGGAGGAACCGCAAAAGAAATGGAAGCGTTTAGCGATGGA	77862
Qy	1400	agctatgtagaagacgtgagatcatcatcaatgaacaatcgaagaatttat	1450
Db	77861	AGC---GGGTAGAGAGGAGGACTTTCGGATAAGAACATGAGGCGTTTGT	77814
RESULT	6		
LOCUS	AC007153/c		
DEFINITION	Arabidopsis thaliana chromosome I BAC F3F20 genomic sequence, complete sequence.	PLN	17-MAY-1999
ACCESSION	AC007153		
VERSION	AC007153.2	GI:4580365	
KEYWORDS	HTG.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1 (bases 1 to 103223) Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, I., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V. S., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 103223) Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, I., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-MAR-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
REFERENCE	3 (bases 1 to 103223) Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, I., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-APR-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
REFERENCE	4 (bases 1 to 103223) Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, I., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-MAY-1999) DNA Sequencing and Technology Center,		





Db	27789	GAAACAGATAGAGGAACCTAGCGAGAGCACTCATAGAAGGGAACACCGTGTGGGT	277330
Qy	989	gataa--gaacaacagaggcgacgaaca-agaacagcagagaagaagaagaagct	1045
Db	27729	TATAACTGATTAATCCAACAGAGAACCAGAAACAGAGGAGAAGAGACAGAGATTGA	276707
Qy	1046	gctgagctcttggatgcgtccaggaactgaacgacctcgggaaaaactgacatggtgctc	1105
Db	27669	GAAGATAGCTGGATTTCACACAGAGCTTGAGAGAGTTGGGATGATTGTGCTGGGTTC	276100
Qy	1106	aaattggatgtcttcgacgcataagtgcggtgggatcgtctcgacgcattgcggttgaaa	1165
Db	27609	GCAGATAGAGGTTTTAAGTCAACCGAGCCGTAGGTTGTTTGTGACTCATTGTGGGTGGAG	275505
Qy	1166	tcttgcatcacagagcctggcttggtggtggtgccctggtggtgcttctcctaagtctga	1225
Db	27549	CTCGACGCTGAGAGGTTTGTTGCTTTGGCGTTGCCGGTTGCGATGTTGTCGATCGGA	274900
Qy	1226	tcaaggagactaatgcgaagatgatcgaagatggtgtagggagttggtgtagagagtcagagt	1285
Db	27489	TCAACCGACGAACCGAAGCTACTTGSAGAAAGTTTGGAGACTGTTGTGAGGGTAGAGA	274303
Qy	1286	gaatgagaaagcggcgttgttgatagcgctgagattaagaaggtgcgtctcggaggttat	1345
Db	27429	GAACAAGGA---TGGTTTGGTGGAGAGAGAGATCAGAGGAGTGTTTGGAAAGCCGTGAT	273737
Qy	1346	-----aaagatcgagagttgagagaagaagccaatgatgtggaaggtttggtcctaaaga	1399
Db	27372	GGAGGAGAAGTCGGTGGAGTTTGAGGGGAACCCAAGAAATGGAAGCGTTTACCGATGGA	273133
Qy	1400	agctatgatatgaagacgtggtatcatcaatgaacaatctgasaattttat	1450
Db	27312	AGC---GGTAGAAGAAGGAGGATCTTCGGATAAAGAACATGGAGGCTTTGT	27265
RESULT	7		
LOCUS	ATCHRIV38	198944 bp	DNA PLN 16-MAR-2000
DEFINITION	Arabidopsis thaliana DNA chromosome 4,	contig fragment No. 38.	
ACCESSION	AL161538		
VERSION	ALI61538.2	GI:7268110	
KEYWORDS			
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophytes;		
	Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; con-		
	eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;		
	Arabidopsis.		
REFERENCE	1 (bases 1 to 198944)		
AUTHORS	EU Arabidopsis sequencing project.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer		
	Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:		
	lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project		
	Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge		
	Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK.		
	E-mail: michael.bevan@bsrc.ac.uk		
COMMENT	Information on performance of analysis and a more detailed		
	annotation of this entry and other sequences of chromosomes 3, 4		
	and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal		
	this fragment has an overlap with ATCHRIV37 at the 5' end and an		
	overlap with ATCHRIV39 at the 3' end.		
FEATURES	Location/Qualifiers		
source	1..198944		
	/organism="Arabidopsis thaliana"		
	/variety="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="4"		
gene	4845..5598		
	/gene="AT4g14060"		
CDS	join(4845..5031,5330..5598)		
	/gene="AT4g14060"		





Intron complement(17605..17686)  
 exon /number=2  
 exon complement(17687..17732)  
 exon /number=3  
 exon 18189..18494  
 /gene="AT4g14090"  
 /number=1  
 gene 18189..25761  
 /gene="AT4g14090"  
 CDS join(18189..18494,18694..18787,18938..19018,19571..19724,  
 19772..19833,21489..21524,22014..22792,22844..22948,  
 23060..23200,23280..23566,24375..24631,24686..25140,  
 25192..25311,25445..25614,25665..25761)  
 /gene="AT4g14090"  
 /note="similarity to proton-coupled peptide transporter  
 PEPT, Rattus norvegicus; Contains Prokaryotic membrane  
 lipoprotein lipid attachment site AA535-545; Prokaryotic  
 membrane lipoprotein lipid attachment site AA955-965;  
 contains EST gb:AI998193.1"  
 Query Match 11.5%; Score 192.6; DB 8; Length 198944;  
 Best Local Similarity 53.2%; Pred. No. 1.6e-36;  
 Matches 750; Conservative 1; Mismatches 550; Indels 108; Gaps 12;  
 Qy 60 catattctactagcaacattccagcacaagccacataaaacctctctctcgagttcgcc 119  
 Db 21208 CATTACTTGTGTAAACATCCAGAGCAGGTACATCAACCGCGCGCTTCAACTAGCC 21149  
 Qy 120 aaaaggtctcaacacccggatacgtcgaccaagtcacattcttcacgagtgatacgca 179  
 Db 21148 AACCGCCTCATCCACCACCGGTGCAACC-----GTCACATATCTCCACCGCAGTCTCTGCT 21095  
 Qy 180 ttgagacgcatgcctctgaaacgcatccgagcagaagaatcgtattctctggtgcatkyca 239  
 Db 21094 CACCGACGTATGGG---CGAGCCACCTTCCACAAAAGTCTATCTTCTGCGTGTGTTACCC 21038  
 Qy 240 gattctacgatgatggcttaagaagggcgacgatggcaaaactacatctcgagatg 299  
 Db 21037 GATGATTTCGACGCGGTCTCAAGTCATTGCAAGCAGCAAAATCTACATCTCCGAATC 20978  
 Qy 300 aagaagcgcgaacgaaggccttaagaagacactcttattaagctcaacogatgcgcgatg 359  
 Db 20977 AAACGATGTGGTTCAAACGCCCTGAGAGACATCATCAAAAGCCAAATCTTGAAGCCACCACC 20918  
 Qy 360 ggaagtgaatttacaatcgcgtgagcttctgtgtactctcatctattcttgagca 419  
 Db 20917 GAAACAGAGCTATCACCGGGTAA-----TCTACTCTGTCTCGTCCCGTGGGTT 20867  
 Qy 420 gctgaagtggcgcgtaagtcgagtcgagtgccctcttcttgattgaaccggctacg 479  
 Db 20866 TCTACGGTAGCGCGTGAGTTTCACCTCCAACTACACTTCTCTGATTGAACCAAGTACT 20807  
 Qy 480 gtttcgatgtgactatttttaactcaatgggtatgccgatgatatacgatcgaggctca 539  
 Db 20806 GTACTAGACATCTACTACTACTTCAACACCTCTTACAAACATCTC---TTCGACGTT 20750  
 Qy 540 gatcaaatccaactgccaatctccagctctcccaagcaagatccctcttctcta 599  
 Db 20749 GAACCGATTAAATACCGAACTGCCACTGTATCACCGGTGACCTCCCTCGTCTTCTT 20690  
 Qy 600 ctcccttcgagccc---cgcgagattccgaaccctaatgaagaaaagttcgacacgctc 656  
 Db 20689 CAACCTTCGAAGGCATTACCGTCACTCTTGTGACTCTAAGAGAACAATATCGAAGTCTC 20630  
 Qy 657 gacaagaacgaagaagcgaaggtttgataaacaacgttcgcagcattagaaacccaacaa 716  
 Db 20629 GAAACGGAATCAAAACCTTAAGATTCCTGTTAAACACATCTCTGCTTTGGAAACAGGATCT 20570  
 Qy 717 ctcaaaacccatcgacaggtatgaactaatatccatcgcccaattaatcccatcatcgata 776  
 Db 20569 TTAACCTCTGTTGAGAACTCAAGATGATCCCAATCGGACCGTTGGTTCTTC----- 20515  
 Qy 777 ttctcagatggcaacgacccctctcatcaagcaaaatccctacggtgagacctcttcaga 836

Db 20516 --CTCCGAGGGTAAAAACC-----GATCTTTTCAA 20489  
 Qy 837 aaagccgatgaacttaccatgactgctaaactcaaaacccgaatcctcggtcgtttac 896  
 Db 20488 TCTTCCGACGAGGATTACACGAAATGGTTAGACTCGAAGCTCGAGAGATGATGATTAC 20429  
 Qy 897 gttcgttcggagcctcctgagg---ctcccgaaccccaatggaagaatagcaata 953  
 Db 20428 ATTTCTTTAGGCACACACGCCGATGATTTACAGAGAAACATCGAAGCCCTTACTCAC 20369  
 Qy 954 gggcttcagacacacccaatcgcctgctctctctgggtgataaagaagaacgaagagggcgac 1013  
 Db 20368 GCGCTGTTAGCTACAAACAGACCGCTTTTATGATCGTGAG----- 20326  
 Qy 1014 gaacaagacagaagaagaagaagctgctgagctctcttcttgatcgctcaggaact 1073  
 Db 20327 -----GGAGAAAAATCCAGAAAGAGAAGAAATCGGTTTCTTGAATTGATCAGAGGA 20273  
 Qy 1074 gaagcactcgggaaaaatcgtagcatggtgcacaaattggatgttctgacgcataaagtcg 1133  
 Db 20272 AGTATCGAGGATTTGGTGGGATGGTGTCTCAGACAGCTGTTTGGCGCATTTGTCT 20213  
 Qy 1134 gtgggactctcgtgacgacattgctggttggaattctctatcgagagcctggtctgtggt 1193  
 Db 20212 GTGGGATGTTTGTGACTCATTTGTTGGAAATTCGAGCTTGAGAGATTAGAGAGTGT 20153  
 Qy 1194 gtccccgtggtgcttctctcctcaatggttcgatacaagggaactaatgcgaagatgacgaa 1253  
 Db 20152 GTTCCGTTGGTGGTTCGCCAGTTTCTGATCAGTGTACACGGCGGAAGCTTGTGGAG 20093  
 Qy 1254 gatgtggaagagtggtgtagagtcagatgaatgaagagggcggtgctgtagatgg 1313  
 Db 20092 GATCGTGGAGATTGGAGTGAAGTGAAGT---TGGGGAGGAAGGAGATGTGGATGG 20036  
 Qy 1314 cgtgagattaagagtgctcgtcgaggttataaagagtcga-----gagttg 1361  
 Db 20035 GAGGAGATTAGAGGTGTTTGGAGAAGCTGATGCTGTGGAAGAGCGCGGAGGATG 19976  
 Qy 1362 agagaagcgcgaatgatgtggaaggggttggcttaagaagctatgtagaagaacgtgga 1421  
 Db 19975 AGAGAGATTCAGAGAGAGTGGAAAGCGATGGCTGTGTATGC---GGCAGCGGAAGTGGA 19919  
 Qy 1422 tcataatgaacaactcgaagaattttat 1450  
 Db 19918 CCGTCCGATTGTAATCTTAAAGGTTTGT 19890  
 RESULT 8  
 ATFCAD/C 200576 bp DNA PLN 28-JUN-1999  
 LOCUS Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment  
 DEFINITION No. 0.  
 ACCESSION 297335  
 VERSION 297335.2 GI:5280985  
 KEYWORDS  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 eudicotyledons; Magnoliophyta; Magnoliopsida; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.  
 REFERENCE 1 (bases 1 to 200576)  
 AUTHORS Bevan, M., Stekema, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N.,  
 Kreis, M., Kavanagh, T., Entian, K.D., Rieger, M., James, R.,  
 Pridmore, P., Hatzipoulos, P., Obermaier, B., Duesterhoft, A.,  
 Jones, J., Palme, K., Ansong, W., Delseny, M., Bancroft, I.,  
 Mewes, H.W., Schueller, C. and Chalhatzis, N.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 200576)  
 AUTHORS EU Arabidopsis sequencing project.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer

Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schuellemips.biochem.mpg.de, mayerremips.biochem.mpg.de  
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,  
 E-mail: michael.bevan@bbsrc.ac.uk  
 On Jun 30, 1999 this sequence version replaced gi:2244747.  
 Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>  
 this fragment has an overlap with AIFCAL at the 3' end.

## FEATURES

source

1..200576

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db\_xref="taxon:3702"

/chromosome="4"

source

1..63289

/organism="Arabidopsis thaliana"

/db\_xref="taxon:3702"

/clone="BAC TAMU20K8"

gene

1753..4428

/gene="dl3000w"

join(1753..3194,3261..4428)

CDS

/gene="dl3000w"

/note="similarity to Cf-2.2, Solanum pimpinellifolium"

/codon\_start=1

/product="disease resistance Cf-2 like protein"

/protein\_id="CAB10171.1"

/db\_xref="GI:2244748"

/translation="MITIWSLCLIFCLSNLSILVIKDLCLPDQDALLKFNFSIP

SPDSMLILQTTAKRNNTDCCSMGGISCDPKTGVVVELDGLNSDLNGLRNSSLP  
 RLQHLQSLDLYNDLSCTLPDSSGNFKYLRVNLGLCNLFGEIPTSLSLYLDL  
 SYNDNTEGLIDSMGNLAKLRVLSITCKFTGKIPSSDGLNLYLFDLDSWYFTEL  
 PDSMGNLRLVNLHRCNFEKIPITSLSLSNLTDLDISKNEFTSEGPSMSLNRL  
 TSLIKDLGTDNFSGLKIGNISSPNLQELYIGENNINGIPRSLIKLVLSLSL  
 FWDGTGIVDFISFLQKLSRLDLSGLNLTSSSHLPSMMHLILSSCNISQPKFL  
 ENQSLYHLDIANQIEQVPEWLRLPTLSFIADNKFSEIIPRANCEITLVLSNN  
 NFSGIPPCFEISNTLSILHRLNSLGSVIPESLHGLKLSLGVNSRLSGQPKSL  
 INCYLQGLVNEENKINDTPSWLKSPLNQLLVLSRNFEGHPIFSPGDSLSFKLRF  
 FDISNRSGLVPLSPDYFGVMSVMSFVDIIDIPTGTVVGGDQSFHKSVVLTIKGLN  
 MELVSGSEIYKTIIDVSGNRLGDIPEISIGILKELIVLMSNNAFTGHIPPSLSLNS  
 LQSLDLSQNLSGISIPGELGTFLARNFYSNMLEGPIPOGTQIQSONSSSEAFENPG  
 LCGAPLOKCGSEEDKEKEKDKGLSWAAAIGYVPLFCGLAIGHILISYKRDWF  
 WRIFSCFSSPLKK"

1753..3194

/gene="dl3000w"

/number=1

3195..3260

/gene="dl3000w"

/number=1

3261..4428

/gene="dl3000w"

/number=2

complement(5905..6149)

/number=1

5905..7913

/gene="dl3005c"

complement(join(5905..6149,6243..6344,6436..6826,

7236..7913))

/note="strong similarity to glycine

hydroxymethyltransferase (EC 2.1.2.1); contains EST

gb:1565775"

/codon\_start=1

/product="hydroxymethyltransferase"

/protein\_id="CAB10172.1"

/db\_xref="GI:2244749"

/db\_xref="SPTREMBL:O23254"

/translation="MEPVSSWNTSLVSDVPEIHDLEKRRQRCGIELIASENFTS  
 FAVIEALSNLTNRYSEGIPGNRYGNEFIDEINLCRSALAEAFHCDPAAGVNVQ  
 PYSGSPANFAVATLQPHRDMGLDPLSGHLTHGYVTSGGKISATSIYFESLPYK  
 VNFTGYIDYDKLEKALDFRPKLLICGSAVPRWDYARPRATADKYAGVALLCDMAH

ISGLVAQEAANPEYCDVVTTHKSLRPRAGMIFYRKGPKPKKQPGGAVYDFE  
 DKINFAPALQGGPHNHOIGALAVALKOANTPGFKVYAKOVKANAVAGNVLMSKGY  
 QIVTNGENHLVLDLRLPLGLTGKNEKVKLDCSTLNKNAVFGDSALAPGVRIGA  
 PAMTSRGLVXKDFEQIGEFELSRVTLTLDIOKTYGKLLKDFNKGLVNNKDLQLKADV  
 EKFSASYEMPGLFMSMKYKD"  
 complement(6150..6242)  
 /number=1  
 complement(6243..6344)  
 /number=2  
 complement(6345..6435)  
 /number=2  
 complement(6436..6826)  
 /number=3  
 complement(6827..7235)  
 /number=3  
 complement(7236..7913)  
 /number=4  
 12823..14568  
 /gene="dl3010w"  
 join(12823..13533,13822..14568)  
 /gene="dl3010w"  
 /note="strong similarity to adenosylhomocysteinase (EC  
 3.3.1.1)"  
 /codon\_start=1  
 /product="adenosylhomocysteinase"  
 /protein\_id="CAB10173.1"  
 /db\_xref="GI:2244750"  
 /db\_xref="SWISS-PROT:O23255"  
 /translation="MALLVQTSRGREYKVKDMSQDGRLELELAEVMPGLMACRT  
 EFGSQPKGARITGSLHMTQTAVLIEITLALGAERWCSCNIFSTQDHAAALARD  
 GAVPAFWKGTLOEYWCETERALDWMGPGGDLIVDDGDATLLIHEGVKAEIEFKT  
 SOVDPSTDNPEFOIVLSIIKEGLQVDPKKYHKMKRLVGVSEITTVGKRLYOMQO  
 NGTLFLPAINVDSVTYSKFDNLYCGRSLDPGLMRATDVMAGRAVLCIGDYGKG  
 CAAAKTAGARIVITEIDPICALQALMEGLQVLTLEVEDVSEADIVFTTGNKDIILWD  
 HMRKNNAIVCNIGHFNEIDMLGETYPGVKRITIKPQTDWRVFPETKAGIIVLAE  
 GRMLKGCATGHPSEVMSCTNOVIAQLELWENKASGKYEKKVYLPKHLDEKVAL  
 HLKGLKCATKLSQSDSVYSIPIEGYKPPHYR"  
 12823..13533  
 /gene="dl3010w"  
 /number=1  
 13534..13821  
 /gene="dl3010w"  
 /number=1  
 13822..14568  
 /gene="dl3010w"  
 /number=2  
 complement(16093..16270)  
 /number=1  
 16093..16501  
 /gene="dl3015c"  
 complement(join(16093..16270,16317..16501))  
 /note="weak similarity to Walleye dermal sarcoma virus gag  
 polyprotein"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="CAB10174.1"  
 /db\_xref="GI:2244751"  
 /db\_xref="SPTREMBL:O23256"  
 /translation="MNSTALVIFAIFLLISADAFPIPSNPEIDAMLIRNSIIGED  
 EDLMPTETSRVRLMAQKRKNQELLTIVALDKLILTAEAATPLPDALETQTTSTHEGKI  
 QETAKIGSFLAASKLYD"  
 complement(16271..16316)  
 /number=1  
 complement(16317..16501)  
 /number=2  
 complement(join(18573..18914,18997..19137,19212..19554,  
 19603..19719,19762..20123))  
 /note="weak similarity to fusion protein la/1b - sugar  
 beet"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="CAB10175.1"  
 /db\_xref="GI:2244752"

```

/db_xref="SPTREMBL:O23257"
/translation="MDHVSLSLDEVLVHILSFTTKEAALTSLSKRWRLFTFVPL
DIDSLVLPQEGKEDRYETQSFMEFVDVRLALQNSPIKLSLKLRTGFDSDRV
WISNALARGVELDLIIILNLVTLKNSVRVWLAAGDIFLPMKILYLSVRLCVDN
VWNGKDVNSASLKITVNYNICIGTFSVDTPSLVYFCFEYVAIDYVPMQNL
EARSILVTDQTERARAPNDIVEDVVLRLVNLKIRGIRNVQMYLSPTNLVL
SLCESMFPNNLSLTLKSHKRGQAMPVLLRNCPLHLETLVLEGTLHHVTKCGYA
CDCVSEKGCSTSCPVKVLEIKFGRGTIKEMHMIKHFVNYFPCLMETNIIYTESGP
PKLRVPQCEFTAEEMEHYKSSSENVQLLVGGLYKKWTAQ"
complement(18573..18914)
/number=1
gene 18573..20123
/feature="di320c"
complement(18915..18996)
/number=1
complement(18997..19137)
/number=2
complement(19138..19211)
/number=2
complement(19212..19554)
/number=3
complement(19555..19602)
/number=3
complement(19603..19719)
/number=4
complement(19720..19761)
/number=4
complement(19762..20123)
/number=5
misc_feature 20514..20606
/feature="EM_ORG:CHSOPSBO Spinach plastid psbB operon with
/feature="EM_ORG:CHSOPSBO Spinach plastid psbB operon with

```

```

Query Match 11.5% Score 192.6; DB 7; Length 200576;
Best Local Similarity 53.2%; Pred. No. 1.6e-36;
Matches 750; Conservative 1; Mismatches 550; Indels 108; Gaps 12;

Qy 60 catattactagcaacattccagcagaagggccacataaaccttctctcgagttcgcc 119
Db 81660 CATTACTTGGTTGTAACATTCCTCCAGCGAAGGTCACATCAACCCGGCGCTTCAACTAGCC 81601

Qy 120 aaaaagctctcaacacggaatacgtcgaccaagtcacattcttcacgagtgatacgca 179
Db 81600 AACCGCCTCATCCACACAGGTGCAAC-----GTCACATATCCACCGCAGTCTCTGCT 81547

Qy 180 ttgagagcagtcgcttcgaaacgcatccgagcagcagaatcgtatttcgtggcaktcyca 239
Db 81546 CACCGACGTATGG-----CGACCGACCTTCCACAAAGGCTATCTTCGCTGGTGTTCACC 81490

Qy 240 gattctacgatgagtgcttaaaagaaaggcgacgatggcaaaactacatgtcgagatg 299
Db 81489 GATGATTTCGACGACGGTCTCAAGTCATTCGAACACAGCAAAATCTACATGTCGCAACTC 81430

Qy 300 aaaaagcgcgaaacgaagggcttaaaagacactcttattaaagctcaacgaatgcgcgatg 359
Db 81429 AAACGATGTGGTTCAACCGCCTCAGAGACATCATCAAAAGCAATCTTACGCCACCAACC 81370

Qy 360 ggaagtaagtattacaatcgcgtgagcttggtggtgactctcattcttcttgaggca 419
Db 81369 GAAACAGACGCTATCACCGGGGTAA-----TCATCTGTGTCGTCGCCGTGGGTT 81319

Qy 420 gctgaagtcgcgcgtgaagtcgagtcgcgagtgccctcttcttgattgaacgcgctacg 479
Db 81318 TCTACGGTAGCGCGTGAGTTTCACCTCCCACTACACTTCTCTGGATTGAACGAGCTACT 81259

Qy 480 gtttcgatgtgactatttttactcaatggggtatgccgagatgatogtcggygctca 539
Db 81258 GTACTAGACATCTACTACTACTTCAACACCTCTTACAAACATCTC---TTGCGAGCTT 81202

Qy 540 gatcaaatccaactgcccattcttcgagctctcccaagaagatctccctcttctcta 599
Db 81201 GAACCGATTAAATTACCGAAATCGCACTGATACACCGGTGACCTCCCGCTGCTTCTT 81142

Qy 600 ctcccttcgagccc---cgagagattccgaacccctaataaagaaagattcgacacgctc 656

```

```

Db 81141 CAACCTTCGAGGCGATTACCGCTGACTCTTGTGACTCTAAGAGAACATATCGAAGCTCTC 81082
Qy 657 gacaaagaacgaagaaggggtcttgataaaacacgtctcgacgcatagaaacgcgaacaa 716
Db 81081 GAAACGGAATCAAAACCCCTAAGATTCTGTTAACACATCTCTGCTTTGGAAACAGGATGCT 81022
Qy 717 ctcaaaagccatcgacaggtatgaactaatccatcgccatcaataaaccatccatcgata 776
Db 81021 TTAACCTCTGTTGAGAAACTCAAGATGATCCCAATCGGACCGTGTGTTCTTC----- 80967
Qy 777 ttctcagatggcaaacgacccctcatcaagcaacaaatcctacggtgggagacctcttcaga 836
Db 80968 -CTCCGAGGGTAAACCC-----GATCTTTTCAAA 80941
Qy 837 aaagccgatgaacttaacatgagtgctaaactcaaaacccgaatcctcggtcggtttac 896
Db 80940 TCTTCCGACGAGGATTACACGAAATGTTAGACTCGAAGCTCGAGAGATGATGATTTAC 80881
Qy 897 gtttcgttcggagacctcctgagg---ctccgaaaccccaaatggaagaaatagcaata 953
Db 80880 ATTCTCTTAGCACACACGCGCGATGATTACAGAGAAACATCGAAGCCCTTACTCAC 80821
Qy 954 gggctttcagacacacaaatcgccagttctctggtggtgataaagaagaacagagggcgac 1013
Db 80820 GCGCTGTTAGCTACAAACAGACCGCTTTTATGATCGTGAG----- 80778
Qy 1014 gaacaagagcagaagaagaagagtgctgagctctcttcttctgacgcataaagtcg 1133
Db 80779 -----GGAGAAAAATCCAGAGAGAGAGAAAGAAATCGGTTTCTTGAATTGATCAGAGGA 80725
Qy 1074 gaagcactcggagaaatcgtagcatggtgtcacaaattggtgtcttcgacgcataaagtcg 1133
Db 80724 AGTATCGAGGATTGGTGGTGGGATGGTGTCTCAGACAGCTGTTTGGCCGATTTGCTGCT 80665
Qy 1134 gtggagatgcttcgtgacgcatggtggttggaaattctgctatcgagagcgtggtgtggt 1193
Db 80664 GTGGGATGTTTGTGACTCATGTTGTTGGAATTCGAGCTTGGAGAGTTTAGAGAGTGGT 80605
Qy 1194 gtgcccgtggtgcttctcctcaatggttcgatacaaggggactaatcgaaagatgatcgaa 1253
Db 80604 GTTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 80545
Qy 1254 gatgtgagagagtggtggtgagagtcagtgatgaaggaagcgcgctgtgtgtagag 1313
Db 80544 GATAGTGGAGAGATTGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 80488
Qy 1314 cgtgagattaaagagtgctcgtcgtcgtgaggttataaagagtcga-----gagttg 1361
Db 80487 GAGGAGATTAGAAGTGTTTGGAGAAAGTGATGATGTTGGAGAAAGAGCGGAGGAGATG 80428
Qy 1362 agagaagcgcgaatgatgtggaaggggttggctaaagaagctatggatgaagaacgtgga 1421
Db 80427 AGAGAGATCGAGAGAGAGTGAAGCGGATGCTGTTGATGC---GGCAGCGAAGGTGGA 80371
Qy 1422 tcatcaatgaacaatctgaagaatttat 1450
Db 80370 CCGTCGGAATTGAATCTTAAAGGTTTTGT 80342

```

```

RESULT 9
LOCUS ATU81293 1589 bp mRNA PLN 02-JUN-1997
DEFINITION Arabidopsis thaliana UDP-glucose:indole-3-acetate
beta-D-glucosyltransferase (Iagu) mRNA, complete cds.
ACCESSION U81293
VERSION U81293.1 GI:2149136
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;

```

Arabidopsis.  
 1 (bases 1 to 1589)  
 Graham, R.A. and Thornburg, R.W.  
 DNA Sequence of UDP Glucose:Indole-3-acetate  
 Beta-D-Glucosyltransferase from Arabidopsis thaliana (Accession No.  
 U81293) (PCR97-044)  
 Plant Physiol. 113, 1004 (1997)  
 2 (bases 1 to 1589)  
 Thornburg, R.W. and Graham, R.A.  
 Direct Submission  
 Submitted (06-DEC-1996) Biochemistry and Biophysics, Iowa State  
 University, Ames, IA 50011, USA  
 Location/Qualifiers  
 1..1589  
 /organism="Arabidopsis thaliana"  
 /variety="Columbia"  
 /db\_xref="taxon:3702"  
 26..1589  
 /gene="laglu"  
 26..1450  
 /gene="laglu"  
 /EC\_number="2.4.1.121"  
 /function="transfers the glucose moiety from UDP-glucose  
 to indole-3-acetic acid to form ester linked IAA-Glc"  
 /note="IAA-glucose synthase"  
 /codon\_start=1  
 /product="UDP-glucose:indole-3-acetate  
 beta-D-glucosyltransferase"  
 /protein\_id="AAB58497.1"  
 /db\_xref="GI:2141927"  
 Translation="MANNSNSPTGPHELFVTFFAQGHINPSLELAKRLAGTISGARV  
 TFAAISAYNRMFSTENVPETLIFATYSDGHDGFKSSAYSDEKSRDQATNFMSEMR  
 RGKFTLELDEKQNPCTVVYITLLTWAELEAFELHPLSALLAVQPVVPSIF  
 YHYFNGYEDATSEMANTESSIKLPSLLVRDIPSFIVSNVYAFLLPAPREDIS  
 LKEEINPKILITDELEPEAMSSVPDNFKIVPVGLLTLKTRDFSSRGEYIEWLDTKA  
 DSSLYVDFGTVLAVLSKKOLVELKALIOSRRPFLWITDKSYRNKEDEQKEEDCIS  
 SSEKGFIDIGVWCDQFVLNHRISIGCFVTHCGWNSLTSLVSGVPVAFPPQWNDQ  
 MTNAKLDECKNTGVRVMEKEEGVVVVSEIIRRCIEEVMEDKAEERFGRNATRWKD  
 LAEAVERGGGSFNLKAFVDEHM"  
 1568..1575  
 /gene="laglu"  
 1588..1589  
 /gene="laglu"  
 /note="24 A nucleotides"  
 BASE COUNT 433 a 342 c 375 g 439 t  
 ORIGIN  
 Query Match 10.0%; Score 166.8; DB 8; Length 1589;  
 Best Local Similarity 50.1%; Pred No. 2.5e-30;  
 Matches 729; Conservative 1; Mismatches 633; Indels 93; Gaps 9;  
 QY 1 aacacataaaaaataaaagaagaataatataaaaaaaatggtaacaaacgcc 60  
 DB 3 AACACCAACCAATAATTAACAAATGCCCAACAAATTCCAACTCTCCACCGGTCCAC 62  
 QY 61 atattctactagcaacattccagcaagaagccacataaaccttctcagattcgcca 120  
 DB 63 ACTTTCTATTCGTAACATTTCCAGCCCAAGGTTCACATCAACCCCATCTCGAGTAGCCA 122  
 QY 121 aaaggctctcaacacggatcagtcgaacagtcacattcttcacagagtgatatacgcat 180  
 DB 123 AAGCCTTCGCCGGAACAATCTCTGGTCTCGAGTCACCTTCGCCGCTCAATCTCTGCCT 182  
 QY 181 tgagacgctgctgcgaacacgattccagcagcagaa---tcgatttcgtggcaktky 237  
 DB 183 ACAACCGCGCATGTTCTCTACAGAAAACGTCGCCGAAACCCCTAAATCTTCGCTACT 242  
 QY 238 cagattctacagatgctgcttaagaaaggcagcagatggcaaaactacatgctcgaga 297  
 DB 243 CCGATGGCCACGACGACGCTTTTCAATCTCTGCTTACTCCGACAAATCTCGTCAAGAGC 302  
 QY 298 tgagaaagcggaagcgcttaaaaggacacttatttaagctcaacgatgctgcga 357

DB 303 CCAC TGGAAACTTTCATGCTGAGATGAGACGAGCTGGCAAGAGACACTAACCGAACTAA 362  
 QY 358 tgggaagtgaattgttacaatcgcgtgagcttt-----gtggtgtactctatctatttt 411  
 DB 363 TCGAAGATAAACGGGAACAAAACAGGCCCTTTTACTTGGCTGGTTTACAGGATTTCTCTCA 422  
 QY 412 cgtggcagctgaagtggcgctgaagtcgacgtgacgtgacgtgacgtgacgtgacgtgacgt 471  
 DB 423 CTGGGTCTGCTGAGCTAGCGCGTGTGAGTTTCATCTTCTCTCTCTCTCTCTCTCTCTCTCT 482  
 QY 472 cggctacggttttcgatgtgtactatttttacttcaatgggtatgocgatgatcatg--- 528  
 DB 483 CAGTAACAGCTCTTCTCCATTTTACCATTACTTCAATGCTACGAAGATGCAATCTCAG 542  
 QY 529 -----atcggggctcagatcaaatccaaactgcccattcttcgagcgtctccaagc 579  
 DB 543 AGATGGCTAATATACCCCTCTAGTTCTTATTTAAATTTACTTCTCTGCCACCTTACTGTCC 602  
 QY 580 aagatctccctcttcttactcctctcag---ccccgcgagatcccgaaacctaatga 636  
 DB 603 GTGATATTCTTCTTTCATGTCTCTTCCATGTCTACGCGTTTCTTCTACCGCGCTTTC 662  
 QY 637 aagaaaaagttcacacgctcgacaaagaaacccgaagcgaaggtcttgataaaacacgttcg 696  
 DB 663 GAGACAGATTGATTCACCTGAAGGAAGAAATAAACCCCTAAGATCCTCATCAACACTTTCC 722  
 QY 697 acgcattagaaacccgaacaaactcaaaagccatcgacaggtatgaactaatatccatcgcc 756  
 DB 723 AAAGGCTTGAAGCAGA-----AGCCATGAGCTCGGTCTCC 756  
 QY 757 cattaatcccatcatcgatcttctcagatggcaagcagccctcatcaagaacaaactcct 816  
 DB 757 AGATAATTTCAAGATTGTCCCTGCTCGCTCGTTTACTTAACGTTGAGAACCGGATTTTCGAG 816  
 QY 817 acggttgagacacctcttcagaaagccgatgaaacttacatggactggcttaaaactcaaaa 876  
 DB 817 TCGCGGTGAA-----TACATAGAGTGTTGGATACTAAAG 851  
 QY 877 ccgaatcatcggtcgctttacgtttctcgaggagcctctcgaggctcccgaaaccccaaa 936  
 DB 852 CGGATTCGTCGTGCTTTATTTTCGTTCCGGGACGCTTCCGCTGTTGAGCAAGAAACAGC 911  
 QY 937 tgagaaataatagcaataggtcttcagacacccaaatcgccaggttctctggtgataagaa 996  
 DB 912 TTGTGGAGCTTTGTAAAGCGTTGATACAAAGTCGGAGACCATTTCTTGTGGGTGATTACGG 971  
 QY 997 gaaacgaagggcgacgaacgaagcagaagcagaagaagaagaagagctgctgagcttct 1056  
 DB 972 ATAAAGTCGTACAGAAATAAAGAGATGAGCAAGAGAGAGAGAGATTCATTAAGTAGTT 1031  
 QY 1057 ttgatcgtcacggaactgaacgactcgggaaaaatcgacatggtggtcacaattgagtg 1116  
 DB 1032 CAGA---GAAGAGCTTCGATGAGATAGGAATGTTGTTTTCATGCTGTGATCAATTTAGGG 1088  
 QY 1117 tcttaacgataaagtcggtggatgctctgaacgcatcggttggaattctgctatcg 1176  
 DB 1089 TTTTGAATCATAGATCGATAGGTGTTCGTGACGATTCGGGGTGAACCTTACCGCTGG 1148  
 QY 1177 agagcctggcttggtgctgcccgtggtgcttcttcccaatggttcgatacgaaggacta 1236  
 DB 1149 AGAGCTTGGTTTCAGGAGTTCGGGTGTCGGCTTCCGCAATGGAATGATCATGACGCA 1208  
 QY 1237 atcggaagatgacgaagatggtggaagagtggtgtgagagtcagagtgaaatgagga--- 1294  
 DB 1209 ACGCGAAGCTTTTGAAGATTTGTTGGAACACAGGTGTAAGAGTGATGGAGAAAGAAAG 1268  
 QY 1295 -----aggcggcgtgtgttgatagcggcgtgagattaaagggtcgctcggaggttat-- 1345  
 DB 1269 AAGAAGGAGTTCGTTGGTGGATAGTGAGGAGATACGGCGGTGCATTTGAGGAAGTTATGG 1328  
 QY 1346 ----aaagagtcgagaggttgagaaagcgaatgatggaaggggttggcttaagaaag 1401

```

Db 1329 AAGACAGCGGAGGAGTTTAGAGAAATGCCACGAGGTGAGGATTTAGCGCGGAGG 1388
Qy 1402 ctatggtgaagaacg 1417
    ||| |||||
Db 1389 CTGTGAGAGAGGAGG 1404

RESULT 10
ATCHRIV42/c
LOCUS ATCHRIV42 195068 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 42.
ACCESSION AL161542
VERSION AL161542.2 GI:7268303
KEYWORDS
SOURCE
    ORGANISM
        Arabidopsis thaliana
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core
        eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
        Arabidopsis.
    1 (bases 1 to 195068)
    EU Arabidopsis sequencing project.
    Direct Submission
    Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
    Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
    lencke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de,Project
    Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
    Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
    E-mail: michael.bevan@bbsrc.ac.uk
    Information on performance of analysis and a more detailed
    annotation of this entry and other sequences of chromosomes 3, 4
    and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
    this fragment has an overlap with ATCHRIV41 at the 5' end and an
    overlap with ATCHRIV43 at the 3' end.
FEATURES
    source
        1. 195068
        /organism="Arabidopsis thaliana"
        /variety="Columbia"
        /db_xref="taxon:3702"
        /chromosome="4"
        /gene="AT4g15560"
        /number=1
    exon
        join(9917..10066,10350..10475,10559..10874,10981..11252,
        11340..11613,11691..11998,12081..12362,12441..12725,
        12813..12953)
        /gene="AT4g15560"
        /note="GT/AC sites ignored to preserve continuity; light
        inducible gene required for chloroplast development;
        Contains prokaryotic membrane lipoprotein lipid attachment
        site AA452-462; Prokaryotic membrane lipoprotein lipid
        attachment site AA591-601; Transketolase signatures
        AA102-121; Transketolase signatures AA505-521; contains EST
        gb:AW004332.1, AF1997953.1, W43562, N96015, W43388"
        /codon_start=1
        /product="DEF (CLAL) protein"
        /protein_id="CAB78598.1"
        /db_xref="GI:7268304"
        /translation="MASSAFAPSYIITKGLSTDCKSTLSRSSRLVTLDPSPCLK
        PNNHSINRRAKYCASLAERGEYISRPPTPLDITNPIHMKNSYKELKQLSDEL
        SDVIFNVSKTGHLGSLGVVELTVALHYIFNTPQDKILMDVGHQSYPHKILTGRGK
        MPTMROTNGLSGFTKRGESEHDCFTGTHSSTTISAGLGMVAGRDGLKGNVAVIGD
        GANTAGQAVEMNAGYLDSDMIVILNDNKKVSLPTATLDGPPPVCAKLSALSRLQS
        NPALREIRVAKGWTQIGPQHQLAKVDEYARGLISGTGSLFELGLYIGPVDG
        KNTDDLVLTKYKSTNTGFLVHVVTGKRGTPYAEARADDYHGVVRFDPATGRQF
        HTTQKQSYITTEALVAEADVDKDVIAHAAGGGTGLNLFQRRFTRCFDVGIAE
        QHVAATGAAGLACGLKPFCAIYSSFMQAYDVQVHDVLDQLKLPFRFAMDRAGLVGAD
        PTHCGADFVTFMACLPNMIVMAPSDPADLPNMVATAVIDRPSCFYPRGNGIGVAL
        PPKNGVPTEIGRGLIKCEGRVALLGYGSVOCCLGAAMLEERGLNVTVADECK
        PLDRALIRSLAKSEHVLITVEEGSIGFGSHVVOFLDGLLGLKLRWPNVLPDRIY
        DGRADQDLAEAGLMEFISHIAATNALNLIGAPREAL"
        9917..12953
        /gene="AT4g15560"

```

```

Intron
    10067..10349
    /gene="AT4g15560"
    /number=1
exon
    10350..10475
    /gene="AT4g15560"
    /number=2
Intron
    10476..10558
    /gene="AT4g15560"
    /number=2
exon
    10559..10874
    /gene="AT4g15560"
    /number=3
Intron
    10875..10980
    /gene="AT4g15560"
    /number=3
exon
    10981..11252
    /gene="AT4g15560"
    /number=4
Intron
    11253..11339
    /gene="AT4g15560"
    /number=4
exon
    11340..11613
    /gene="AT4g15560"
    /number=5
Intron
    11614..11690
    /gene="AT4g15560"
    /number=5
exon
    11691..11998
    /gene="AT4g15560"
    /number=6
Intron
    11999..12080
    /gene="AT4g15560"
    /number=6
exon
    12081..12362
    /gene="AT4g15560"
    /number=7
Intron
    12363..12440
    /gene="AT4g15560"
    /number=7
exon
    12441..12725
    /gene="AT4g15560"
    /number=8
Intron
    12726..12812
    /gene="AT4g15560"
    /number=8
exon
    12813..12953
    /gene="AT4g15560"
    /number=9
repeat_region
    14413..14436
    20338..24557
    /gene="AT4g15570"
exon
    20338..20373
    /gene="AT4g15570"
    /number=1
CDS
    join(<20338..20373,20767..20852,20962..21012,21256..21367,
    21467..21572,21653..21750,21989..22117,22182..22274,
    22625..22678,22957..23015,23122..23206,23286..23765,
    24183..24269,24366..24557)
    /gene="AT4g15570"
    /note="similarity to SEN1 protein, yeast; Contains
    ATP/GTP-binding site motif A (p-loop) AA13-20; contains
    EST gb:N65040"
    /codon_start=1
    /product="SEN1 like protein"
    /protein_id="CAB78599.1"
    /db_xref="GI:7268305"
    /translation="VGLSRKSFVLIQPPGPGTKTOTLSILGAIMHATPARVOSKGTD
    HEYKRGQMTIOEKYNHWRGASFWILGVNPDALIMPDDGDFPTSGNELKPVVNA
    SRKYRVLVCAPSNSALDEIVLRLSSGLRDENAOQYTPKIVRIGLKAHHSVASVL
    DLKXRGSAIDKQAGTGTDTDSIRTAILEEAALVWGFNFRINFAILSFGSGL
    LANSRGFDVVIIDEAAQAVEPATLIPLATRCQVFLVGDGPKQLPATVISTVAQDSG
    GTSNFERLQKQAPYVKMLQYRMHPELIPISQTKPIVIVSLILTPMIASYFTLQITF
    LTENQIRSFPSKQFYEGALEDGSIDIEAQTTTRDWHKRYRCFGPFCFFDIHEGKESQHPGA

```









Db	47120	CGATTTCGTGCTGCTTTATGTTTCGTTCCGGACGCTTGCCGTTGAGCAAGAAACAGC	47061		exon	/gene="AT4g15240" complement(21638..22378) /number=1
QY	937	tgaagaataatagcaatagggcttcacagacacacaaatccagcttctctcgtggtataagaa	996		CDS	complement(Join(21638..22378,22677..23498)) /note="similarity to hypothetical 66.0 kd protein F47D12.7; contains EST gb:AN712915"
Db	47060	TTGTGGAGCTTTTAAGCGTTGATACAAAGCTCGGACACCATCTTCTGTGGGTGATTACGG	47001			/codon_start=1 /product="hypothetical protein" /protein_id="CAB78566.1" /db_xref="GI:7268271"
QY	997	gaacgaagagggcgacgaagagcaagcagcaagaagaagaagaagctcgtagctctct	1056			/translation="MNQTRKRPPDSSLRLDLSLSEPMLEFPAPKPSLTSRIRNIFLL LIFCFIYIFSYGTFNFRREQISSIARSLSVFSRRHLLFSLAASHDSWLRSSYVR LWYSPSTRVAVFLDRGQESDLTLPVIVSKDVSPFPYFPGGLASARVARVAKET VDRGDKDVRWFVFGDDDTVFVDNLVTLSKYDHRKWFYVGSNSEFYDQNVYSFDMA FGGGFAISASLAKVLAKVLDSCLMRYSHMYGSDSRIFSCVAELGYTLTHEPGFHQID VRNITFGLLCARPLSLVLHLHLDADVDFPPKRNRTESVAHLIGAASFDGRLQDSV CDYSLATVTVSVVMGAVQYEGNKLLPDLTLQKTFSTWRGSGVQSNYMFSTREYP RPGCRPLVFFLDSVSGDTEGTCTSYNINHRVGHCHRAEVAERLERIRVLSPLERNV EOVGFTINDRFSVKPVSILPGTFFFGFLKIKQMGLPRRQCCDISPYNKSVMINIR OCMPDELIAMNT"
Db	47000	ATAAGTCGTACAGAAATAAAGAGATGAGCAAGAGAAGGAAGATTGCTAAGTAGTT	46941		Intron	complement(22379..22676) /number=1
QY	1057	ttgatgctacgaagaatgaacgactcgggaaaaatcgtgacatggtgctcacaattggatg	1116		exon	complement(22677..23498) /number=2
Db	46940	T---CAGAGAAGAGCTCGATGAGATAGGAATGGTGGTTTCATGCTGTGATCAGTTAGGG	46884		gene	/gene="AT4g15250" complement(Join(28265..28651,28738..29226)) /note="similarity to CONSTANS protein, Arabidopsis thaliana"
QY	1117	ttctgacgaataagtcgggtgggagcttcgtgacgcatcgtgattggaattcgtctatcg	1176		CDS	/codon_start=1 /product="CONSTANS like protein" /protein_id="CAB78567.1" /db_xref="GI:7268272"
Db	46883	TTTTGAATCATAGATCATAGGTGTTGTTCTGACGCATTCCGGGTGGAATCTACGCTGG	46824			/translation="MEARCDFCGTEKALYCKSDSAKLCLNCVDVNVHSANPLSQRTSR SLICEKSLQPTAVHCMENVSILCCQCMQWASNTCTGLGHRLLQSLNIPYSDCPSDFGK IWSTSLPSTVSLVSPFSDTLQELDDWNGSSVVTQTNLADYSFFPEMNLPKV IEECSGLDCEGINLDAPLNFNASNDIIGSSLDNTKCYEYEDSFKEENNLGFLSL LLPLTSGNVVPMNSLSMSNLTGESNATDQDCGISPGFLIGDSPWESNVSPNKLRL DEAKRYKOKKSKRM"
QY	1177	agagcctggcttgggtgtgcccgtgggtgtgcttctcctcaatggttcgataaggggacta	1236		exon	complement(28265..28651) /number=1
Db	46823	AGAGCTTGGTTTCAGGAGTTCGGGTGGTGGCTTTCGCGAATGGAATGATCAGATGATGA	46764		Intron	complement(28652..28737) /number=1
QY	1237	atcggaagatgatcgaaatcgtgtgagaggagtgtgtgagagtcagagtgaaagagaa--	1294		exon	complement(28738..29226) /number=2
Db	46763	ACGCGAAGCTTTTAGAAGATGTTTGAAACACAGCTGTAGAGATGATGGAGAAGGAAG	46704		exon	31839..31950 /gene="AT4g15260" /number=1
QY	1295	-----agcgcggtgtgtgtagcgtgagattgaagagtcgctcgaaggttat--	1345		gene	Join(31839..31950,32018..33264) /gene="AT4g15260"
Db	46703	AAGAAGGATTTGGTGGTGGATGATGAGGAGATACGGCGTGCATTTGAGGAAGTTATGG	46644		CDS	/note="strong similarity to UTP-glucose glucosyltransferase; Contains UDP-glucosyltransferases signature AA319-362; contains EST gb:T20820, T46737, AI993247.1, AI994260.1, AI996826.1, AI997955.1, T76043, AA720097, AI997288.1, W43072"
QY	1346	----aaagatcgagagttgagagaagcgaatgatgtggaagggttgctgctaaagaag	1401			/codon_start=1 /product="glucosyltransferase" /protein_id="CAB78568.1" /db_xref="GI:7268273"
Db	46643	AAGACAAGCGGAGGAGGACTTTAGAGGAATCCACGAGGTGGAAGGATTTAGCGCGGAGG	46584			/translation="MKKIELVFPSPGIGHRLPTVKLAKOLIGSEHFFNHQSDRLRL YEPISVEKQPPANTLPESVYIEKQKQVDRDVTSRINPTRLAGVFDVDFMFCSSMIDI ANFEGVPCVNIYTSNATFLGILHVQEMDDKDYDSDLDSEVNELEFCLTTPYPVK CLPHIUSSKDWLPFFAAQGRSPRKMGIILVVAEELPHALMKNFNVLDLPQATIPGVP LHDNDGDDDEKRLKLELDRLMDQPPKSVILFCFGSMGGTTEQTRVAVNALNSGHRF LWSLRASPNIMMERPGYKNLEVLDPGFLERTLDKRGVIGHWAPVLEKPAIGGF VTHCGWNSMLESWFQVPMVTPLYAEQKNAPENVEELGLAVEIKRCISGDLILLIGE MEIVTAEDIERAIRCYNEQSDSDVRSRVKEMAEKCHVALMDGGSSKTKALQKFIQDVLEN VA"
QY	1402	ctatgagatgaagaacg	1417		Intron	31951..32017 /gene="AT4g15260" /number=1
Db	46583	CTGTGAGAGAAGGAGG	46568			
RESULT	12					
LOCUS	ATCHRIV41	197419 bp	DNA	PLN		
DEFINITION	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 41.					
ACCESSION	AL161541					
VERSION	AL161541.2	GI:7268270				
KEYWORDS	thale cress.					
SOURCE	Arabidopsis thaliana					
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.					
REFERENCE	1 (bases 1 to 197419)					
AUTHORS	EU Arabidopsis sequencing, project.					
TITLE	Direct Submission					
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mps.blochem.mpg.de, mayer@mps.blochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk					
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.blochem.mpg.de/proj/thal/">http://www.mips.blochem.mpg.de/proj/thal/</a> this fragment has an overlap with ATCHRIV40 at the 5' end and an overlap with ATCHRIV42 at the 3' end.					
FEATURES	Location/Qualifiers					
source	1..197419 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 21638..23498					
gene						



Db 196723 CAAGAGCTTGAGCGACA-----AGCCATGAGTCGGTTC 196690  
Qy 756 ccattaatccatcatgatattctcagatgagcgcaaacccctcatcaagcaacaatcc 815  
Db 196689 CAGATAATTTTCAAGATTGTCCTGTCGGTTCGTTACTAAAGCTTGAGAACGATTTTTCGA 196630  
Qy 816 tacgggtgagacccctcttcagaaaagccogataaaacttacatggactggctaaactcaaaa 875  
Db 196629 GTCGCGGTGAA-----TACATAGAGTGGTTGGATCTAAA 196595  
Qy 876 ccgaatacatcgtgctgttctacgtttcgttggagagccctcctcaggtcccccgaaccccaa 935  
Db 196594 GCGGATTGCTGTGCTGCTTATGTTTCGTTCCGTCGGACGCTGCCGTGTGAGCAAGAACAG 196535  
Qy 936 atggaagaataagcaatagggctttcagacacacaaatcgccagcttctcgtggataaga 995  
Db 196534 CTGTGTGAGCTTGTAAAGCGTTGATACAAAGTCGGAGACCATCTCTGTGGGTGATTACG 196475  
Qy 996 agaaacgaagagggcgacgaacagagcaagcagaagaagaagaagagctgctgagcttc 1055  
Db 196474 GATAAGTCGTACAGAAATAAAGAAGATGAGCAAGAGAAGGAAGATTCATAGGTAGT 196415  
Qy 1056 ttgatcgtcacgaagactgaacactcgggaataatcgtgacatggtgctcacaaattggtat 1115  
Db 196414 TT---CAGAGAAGAGCTCGATGAGATAGGAATGGTGTTCATCGTGTGATCAGCTTTAGG 196358  
Qy 1116 gtctgacgcataagtcggtggtggtgctctcgtgacgcattcgggtggaattctgctatc 1175  
Db 196357 GTTTTGAATCATAGATCATAGTTGTTTCTGTGACGATTCGGGTGGAATCTTACCGCTG 196298  
Qy 1176 gagagcctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1235  
Db 196297 GAGAGCTTGTGTTTCAGGAGTTCGGGTGTCGGGTTCGCCAATGGAATGATCATGATGATG 196238  
Qy 1236 aatgcgaagatgacgaagatggtgagagagtggtggtgagagtcagagtgaaatgagga- 1294  
Db 196237 AACCGAAGCTTTTGAAGATGTTTGTGGAACACAGGTGAAGAGTGAAGAGGAAAGAGAA 196178  
Qy 1295 -----agcgcgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtt- 1345  
Db 196177 GAGCAAGAGTGTGGTGGGTGATGATGAGAGATACCGCGGTGCATTTGAGGAAGTGTATG 196118  
Qy 1346 -----aaagagtcgagagtgagagaagcgcaatgatggtggaaggtgtggtggtggtggtggt 1400  
Db 196117 GAAGACAAGCGGAGGAGTTAGAGGAATGCCACGAGGTGGAAGGATTTAGCGCGGAG 196058  
Qy 1401 gctatgatatgaagacg 1417  
Db 196057 GCTGTGAGAGAAGGAGG 196041  
  
RESULT 13  
AB000623  
LOCUS AB000623 1594 bp mRNA PLN 05-FEB-1999  
DEFINITION Nicotiana tabacum mRNA for glucosyl transferase, complete cds.  
ACCESSION AB000623  
VERSION AB000623.1 GI:1805358  
KEYWORDS JIGT; glucosyl transferase.  
SOURCE Nicotiana tabacum (strain:BY-2) suspension culture cdna to mRNA.  
ORGANISM Nicotiana tabacum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Asteridae; Solanales; Solanaceae; Nicotiana.  
1 (bases 1 to 1594)  
Kojima,H.  
DIRECT SUBMISSION  
TITLE Submitted (23-JAN-1997) to the DDBJ/EMBL/GenBank databases. Hisae  
AUTHORS Kojima, school of agriculture,nagoya university, laboratory of  
JOURNAL Biochemistry; nagoya, furo-cho,chikusa 464-01, Japan  
(E-mail:hisaek@nuagrl.agr.nagoya-u.ac.jp, Tel:052-789-4098,  
Fax:052-789-4094)  
2 (bases 1 to 1594)  
Kojima,H., Hashizume,K., Imanishi,S. and Nakamura,K.

Jasmonate-induced potential glucosyltransferase from tobacco  
suspension cell  
Unpublished (1997)  
FEATURES  
source location/Qualifiers  
1..1594  
/organism="Nicotiana tabacum"  
/strain="BY-2"  
/db\_xref="taxon:4097"  
/tissue\_type="suspension culture"  
52..1455  
/gene="JIGT"  
52..1455  
/gene="JIGT"  
/standard\_name="jasmonate-induced potential  
glucosyltransferase"  
/codon\_start=1  
/product="glucosyl transferase"  
/protein\_id="BAAL19155.1"  
/db\_xref="GI:1805359"  
/translation="MENLNKECHVLIALFPGQHINPSLQFSKKLINLGVKVTLSSSL  
SAPNRILNPKIEGLTFAPSDCYDGNFKGSFDDYHLFNSAIKSHGSERTANLIKSKA  
KNYPTFRVITYTILMDWAGSVAKKLIHPSTLEWIQATVDFDIYYRFTNANFLVNYD  
SQDIIELPLPSLSSDFPSFVDDVKNDAWAVEIKKOIEILNREFTNANFLVNYD  
AULENALRVLKNVTYIGIPLIPSSFLDEKDRKNFPAADMESENKYMELDARANK  
SVIYAFGVAEISSQMBEISQGLKCGRPFLWVIRETLNGEKPEBKTKCKLEKI  
GRIVIGVSNVANKESIIRDEFQKCIETVWGDAEEGELRKNQAKWKDLAKESTKENS  
SNVNLKAYNECLLGHG"  
polyA\_site 1594  
/note="35 A nucleotides"  
BASE COUNT 514 a 238 c 325 g 517 t  
ORIGIN  
  
Query Match 7.9%; Score 131.4; DB 7; Length 1594;  
Best Local Similarity 49.0%; Pred. No. 1.1e-21;  
Matches 706; Conservative 0; Mismatches 651; Indels 84; Gaps 10;  
  
Qy 35 aaaaaaataatggttaacaaacccatttcttactagcaacattccacagacagcca 94  
Db 51 AATGAAACCTTGAAATGAATGCCATGTTCTATTGTCATATTCCAGGCCAAGGTCA 110  
Qy 95 cataaaacccctctcgtggtcgccaaaggctcctcaacacgggatacgtcgaccaga 154  
Db 111 TATTATCCATCCCTTCANITTTCCAAAAAATTAATCAATTTAGT-----GTGAAAGT 164  
Qy 155 cacattctcagagtgatatacgcattgagacgcgcatcgcttcgaaacccgacgcag 214  
Db 165 CACTTTATCTCTAGCCTATCAGCTTTCAATCGCATAAAAATCTCCAAAAATAGAAG 224  
Qy 215 cagaatcgattcgtggcatktycagattcttcaagatgaggtggttaagaagcgagca 274  
Db 225 ATTAACCTTTGCTCCTTCTCTGATGGCTATGATGGAATTTCAAGGTTTCATTGTGATGA 284  
Qy 275 tggcaaaactacatcgtcgagatgagaagcgcgcggaacgccttaaggacactct 334  
Db 285 TTACCAATTTATTCATTCAGCAATAAAATCTCATGGTTC---TGAATTTATGGCAATTT 341  
Qy 335 tattaagctcaacgagtgctgcgatggaagtgaaatgtatacaatcgctgagcttctggt 394  
Db 342 AATTAATCCCAAGG-----CAAAAAATGGTTTACCCCTTTACACGTTGTTAT 386  
Qy 395 gtaactcatctatttctggtggcgagctggaagtcggtgagctgagctgagcgagtcgc 454  
Db 387 TTACATATTCTTATGGAATTTGGGCTGGTTCTAGTAGCCAAAAAATATACATATCCCTTCTAC 446  
Qy 455 cctctcttgatgaacgcggtacgtgttctcgatgtgtactatttttacttcaatgggta 514  
Db 447 ATTATTTGGATTCAACAGCTACAGTTTGTGATATTATTATTATTCGGTTTACTAATTT 506  
Qy 515 tgcgatgatcatcgatcggtggtcagatcaaaatccaaactgcacaaatcttcgcagctctc 574  
Db 507 TGCTAATTTATTTCAAGAATATTGATTTCTCAGGATCAAAATTTATTGAGTACACAGGTTTGC 566

```

QY 575 -----caagcaagatctccctcttctactcccttcgaagcccgagattc 623
Db 567 TTCACCTAGTCAAGTGATTTTCCCTTCATTTGTTGATGATGCAAGCTCAATGATG 626
QY 624 cgaaccccaatgaagaaag-----ttcgacacgctcgacaagaacccaagcgaaggt 679
Db 627 GCAGTTGAGTCAATTAAGGCAAAATTCAGATATTGAATAGTGAAGAAATCCAAGAT 686
QY 680 ctgtataaacgcttcgacgattagaaccgaacaaactcaagccatcgacaggtatga 739
Db 687 TCTTGTGAATACCTTTGATGCAATGGAACCTTAATGCTTTGAGGGTTTGAAGAAATGTAAC 746
QY 740 actaatatcatcgcgccattatcccatcatcgattctctcagatggaacccatcattga 859
Db 800 GAAAGATAATTTTTCGGCGCTGATA-----TGATCGAAAGTGAAATAATACATGGA 854
QY 860 ctgggtataactcaaacccgaatcctggtctgttctacgtttctcggtcgagcctcctgag 919
Db 855 GTGGTGGATGCAAGGGCGCAATTAATACGTTATTACATAGCATTTGTTAGCTATGCTGA 914
QY 920 gctccgcaaaccccaaatggaagaaatagcaataggccttcagacacccaatcgcaggt 979
Db 915 GATATCAAGTCAATGATGGAAGAGATTAAGTCAAGGATTGTTGAATGTGGAAGACCAT 974
QY 980 tctctgggtgataagaagaaacgaagagggcgacgaacgaagcgaagcgaagaaaga 1039
Db 975 TTTGTGGGTGATTAGAGAAACCTTAATGCGGCAAAACCCGAGAA----- 1020
QY 1040 gaagctgctgactcttcttgatcgtcaagaaactgaacactcgcgaaatcgtgacatg 1099
Db 1021 -----AAGTTGACCTGTAAAGATGAATGGAAGAGATAGGAGAAATAGTCGGTTG 1070
QY 1100 gtgctcacaattgattctcgaacataaagtcgtggtggtggtggtggttctcctcaatg 1159
Db 1071 GTGTCGCGAAATGGAAGTCTTAACACTCGTCGCGTAGCTGTTTCTTCACTACTGTGG 1130
QY 1160 ttggaaattctgctacgagcctggtggtggtggtggtggtggtggttctcctcaatg 1219
Db 1131 ATGGAATTCGACTCTGGAGAGCTAGCTTCTGTTGTCGCAATCGTGGCATGCCGATTTG 1190
QY 1220 gtctgatcaaggaactaaatcgaagatgatcgaagatggtggtgaggtggtgagatg 1279
Db 1191 GAATGATCAAAATTTGCAACGCGAAGCTCATCAAGATGTTTGAAGATGTTGAGATGAGAGT 1250
QY 1280 cagagtgaatggaagcggtggtt-----gttgataggctgagattgaaggtgct 1333
Db 1251 GAACGCTAATAAGAGAGGTATCATTAAGAGACGAGCTTCCAGAGCTGTATGAGATCGT 1310
QY 1334 ctccgaggttataaagtcgagag---ttgagagaaacgcaatgatggaaggttt 1390
Db 1311 AATGGAGATGCTGAGGAGAGAGGAATTTGAGAAAGATGCCCAAAATGGAAGATTT 1370
QY 1391 ggtcaagaagctatggtatgaagaacgtggtatcaatgaacaatctgaagaatttat 1450
Db 1371 GGTGAAGGAATACGAAGGAAATAGTTCAATCAATGATAATGTAATGTAATGTAATGTAAT 1430
QY 1451 t 1451
Db 1431 t 1431

RESULT 14
AF190634 1440 bp mRNA PLN 02-APR-2000
DEFINITION Nicotiana tabacum UDP-glucose:salicylic acid glucosyltransferase
(SA-GTase) mRNA, complete cds.
ACCESSION AF190634
VERSION AF190634.1 GI:7385016

```

```

KEYWORDS common tobacco.
SOURCE Nicotiana tabacum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
Nicotiana.
REFERENCE 1 (bases 1 to 1440)
AUTHORS Lee, H.-I. and Raskin, I.
TITLE Purification, cloning, and expression of a pathogen-inducible
UDP-glucose:salicylic acid glucosyltransferase from tobacco
Unpublished
JOURNAL 2 (bases 1 to 1440)
AUTHORS Lee, H.-I. and Raskin, I.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) Biotech Center, Rutgers University, Foran
Hall/Cook College, 59 Dudley Rd, New Brunswick, NJ 08901, USA
FEATURES
Source
1..1440
/organism="Nicotiana tabacum"
/cultivar="Xanthi-nc"
/db_xref="taxon:4097"
1..1440
/gene="SA-GTase"
34..1413
/gene="SA-GTase"
/codon_start=1
/product="UDP-glucose:salicylic acid glucosyltransferase"
/protein_id="AAF61647.1"
/db_xref="GI:7385017"
translation="MTQKAHCLILPYPAOGHINPMLQFSKRSLOSGVKITAAATKSF
LKTQELSTSVSVEAISDYGDDGREGAGTFVAYLTRFEKVGSDLSQLIKLTNCGC
PVSCLVDFPLPMAVEVGNRNVATAAFQSCAVDNIYHVHKGVLKLPFDVDRKEI
SIPGLTIEADSPVSVNPESSRIEMLVNSLENTDWLINSFVELEKEVIDMM
AKIYPTKGTPTSPMYLDRPLDDKEYGLSVKPMTNACLNHLNHPVSVVYVSFG
SLAKLEAQMEELANGLSNKNFLWVRSTESKLPNNFLEELASEKGLVVSQPOL
QVLEHKSIGCFTHCGWNSTLEAISLGVPMIAMPWSDQPTNAKLVEDWVWEMGIRPKQ
DEGLVREVEIEECIKIVNEKKKIRENAKKWELARKAVDEGSSDRNIEEFVSK
LVTIASVES"
BASE COUNT 484 a 239 c 318 g 399 t
ORIGIN
Query Match 6.9%; Score 116; DB 49; Length 1440;
Best Local Similarity 47.9%; Pred. No. 6.6e-18;
Matches 538; Conservative 0; Mismatches 535; Indels 51; Gaps 5;
QY 353 tgcgctggaagtgaaatgttacaatcgctgagctgttggtggtgctctcatctatttc 412
Db 315 TCGAAAGTTTAACAAAATTTGGTGTGCTGCTGAGTTCATAGTTAGATCCATTTCTTC 374
QY 413 gtgggcagctgaagtcgctggaagtcgacgtgcgagtcgctctcttgggattgaacc 472
Db 375 TTGGGCTGTGAAGTGGGAAATAATTTTGAGTAGCTACTGCTGCTTTTTCACACATC 434
QY 473 ggcctagcgttttcgagtgatgctatttcttcaatgggtatggcgatgatacgtatgc 532
Db 435 TTGTGCAGTGGATAACATTTATTACCATGTACATAAAGGGGTTCTTAAACTTCTCTCAAC 494
QY 533 gggctcagat---caaatcaactgcccaatcttcgcgagctctccaagcaagatctccc 589
Db 495 TGAGCTGTATAAAGAAATCTCAATTCCTGGATTATTAAACAATTTAGGCATCAATGTACC 554
QY 590 ctcttctactccttcgcagcccgagattccgaaacccctaatgaagaaagttcga 649
Db 555 TAGTTTGTGTTCTTAATCCTGAATCTTCAAGAAATACTTGAAATGTTGGTGAATCAGTTCTC 614
QY 650 cagctcgcaagaacccaagcgaaggtctcttgataaacaagcttcgacgcatagaac 709
Db 615 GAATCTTTGGAAC-----ACAGATTGGGTCTTAATCAACAGATTTCTATGAATTGGAGA 668
QY 710 cgaacaactcaagccatcgacaggtatgaactaatatccatcgcccatatccatc 769

```

Db	669	AGAAGTAATTCATGGATGGCCACAGATCTATCCATCAAGACAATTTGGACCAACTATATACC	728			
Qy	770	atcgatatcttcagatgggcaacgacccctcatcaagacacaaatcctcaggtggagacct	829			
Db	729	ATCAATGTACCTAG-----ACAAGAGCTACCCAGATGACAAGAATATGGCTTAGTGT	782			
Qy	830	cttcagaaaaaccgatgaaacttacatggactggctggttaactcaaaacccgaatcatcggt	889			
Db	783	CTTCAAGCCAATGACAAATGATGCTTAACCTGGTTAAACCATCAACCAAGTTAGTCTAGT	842			
Qy	890	cggttaagttcgttcggagccctcctgaggtcccgaaaccccaatggaagaaataacg	949			
Db	843	AGTATATGATCATTTGGAGTTTGTAGCCAAATAGAAAGCAGACAAATGGAAGAATAGC	902			
Qy	950	aatagggtcttcagacacaaatcccgagttctcgtggtgaataagaaagaaacgaagaggg	1009			
Db	903	ATGGGTTTGAGTAATACCAACAACAACCTCTCTGTGGTAGTTAG-----	947			
Qy	1010	cgacgaacagagcaagcagaaagaaagaaagctgctgagctctcttctgctgcgcacgg	1069			
Db	948	-----ATCCACTGAAGAATCCAAACTTCCCAACAACCTTTTATGAGGAAT	992			
Qy	1070	aactgaacgactcgggaaaaatcgtagcatggtctcaaatctggatggtctcagcgataa	1129			
Db	993	AGCAAGTGAAAAGGATTTAGTCGTCTCATGTTGCCAATTTACAAAGTCTTGGAAACATAA	1052			
Qy	1130	gtcgggtggatgcttcgtagcattcggttggaattctgctatcgagagcctgctg	1189			
Db	1053	ATCAATAGGGTGTTCACGCACTGTGGCTGGAAATCAACTTTGGAAGCAATTTAGTTT	1112			
Qy	1190	tgggtgcccgtggtgctcttcctcaatggtctgatcaagggagctaatcgaaagatgat	1249			
Db	1113	GGGAGTACCAATGATGTGAATGCCACATTTGGTCAGACCAGCCACAATTCGAGGCTTGT	1172			
Qy	1250	cgagatgctgaggagagctggtgtagagctcagatgtaagtagggagggcggcgtt	1304			
Db	1173	GGAAAGTCTTTGGGAGATGGGAATTTAGACCACAAAGATGAAAAGGATTTAGTTAGAAG	1232			
Qy	1305	-gttgatagcgtagatgaagagctcgctcggaggtataaagagtcgagagttgag	1363			
Db	1233	AGAAGTTATTCAAGANTGATTTAAGATAGTATGGGAGGAAAGAAAGAAAGATTAG	1292			
Qy	1364	agaagcgcaatgatgtggaaggggttggctaaagaagctatggatgaagaacgtggatc	1423			
Db	1293	GGAAATGCAAAAGAAATGGAAGGAATTTGGCTAGGAAGAGCTGGATGAAGGAGGAAGTTC	1352			
Qy	1424	atcaatgaacaatctgagaatttttactactaggtattaatg	1467			
Db	1353	AGATAGAANAATTTGAAGAATTTGTTTCCAAAGTTGGTGACTATTG	1396			
RESULT 15						
AB019232						
LOCUS	AB019232	80818 bp	DNA			
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MIL23, complete sequence.					
ACCESSION	AB019232					
VERSION	AB019232.1	GI:3869071				
KEYWORDS	HYG.					
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1 clone:MIL23.					
ORGANISM						
Arabidopsis thaliana						
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.						
1 (sites)						
Nakamura.Y.						
Structural Analysis of Arabidopsis thaliana Chromosome 3. I						
Unpublished (1998)						
2 (bases 1 to 80818)						
Nakamura.Y.						
Direct Submission						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						

[illegible]

Search completed: August 1, 2000, 13:21:41  
Job time: 56763 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 14:38:28 ; Search time 3616.02 seconds  
(without alignments)  
2037.827 Million cell updates/sec

Title: US-09-147-955-7  
Perfect score: 1671  
Sequence: 1 aacacataaaaaaaaaataa.....ctaaaaaaaaaaaaaaaaaa 1671

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues  
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	gb_est1:*
21:	gb_est2:*
22:	gb_est3:*
23:	gb_est4:*
24:	gb_est5:*
25:	gb_est6:*
26:	gb_est7:*
27:	gb_est8:*
28:	gb_est9:*
29:	gb_est10:*
30:	gb_est11:*
31:	gb_est12:*
32:	gb_est13:*
33:	gb_est14:*
34:	gb_est15:*
35:	gb_est16:*
36:	gb_est17:*
37:	gb_est18:*
38:	gb_est19:*
39:	gb_est20:*
40:	gb_est21:*
41:	gb_est22:*
42:	gb_est23:*
43:	gb_est24:*
44:	gb_est25:*
45:	gb_est26:*
46:	gb_est27:*
47:	gb_est28:*
48:	gb_est29:*
49:	gb_est30:*
50:	gb_est31:*
51:	gb_est32:*
52:	em_est20:*
53:	em_est21:*
54:	em_est22:*
55:	em_est23:*
56:	em_est24:*
57:	em_est25:*
58:	em_est26:*
59:	gb_est33:*
60:	gb_est34:*
61:	gb_est35:*
62:	gb_est36:*
63:	gb_est37:*
64:	gb_est38:*
65:	em_est27:*
66:	em_est28:*
67:	em_est29:*
68:	em_est30:*
69:	gb_est39:*
70:	gb_est40:*
71:	gb_est41:*
72:	gb_est42:*
73:	gb_est43:*
74:	gb_est44:*
75:	em_est31:*
76:	em_est32:*
77:	em_est33:*
78:	em_est34:*
79:	gb_est45:*
80:	gb_est46:*
81:	gb_est47:*
82:	em_est35:*
83:	em_est36:*
84:	em_est37:*
85:	gb_est48:*
86:	gb_est49:*
87:	gb_est50:*
88:	gb_est51:*
89:	gb_est52:*
90:	gb_est53:*
91:	gb_est54:*
92:	gb_est55:*
93:	gb_gss1:*
94:	gb_gss2:*
95:	gb_gss3:*
96:	gb_gss4:*
97:	em_gss1:*
98:	em_gss2:*
99:	em_gss3:*
100:	em_gss4:*
101:	gb_gss5:*
102:	gb_gss6:*
103:	gb_gss7:*
104:	gb_gss8:*
105:	gb_gss9:*
106:	em_gss5:*
107:	em_gss6:*
108:	em_gss7:*
109:	em_gss8:*
110:	em_gss9:*
111:	em_gss10:*
112:	em_gss11:*
113:	gb_gss10:*
114:	gb_gss11:*
115:	em_gss12:*
116:	gb_gss12:*

117: gb\_gss13:\*  
118: gb\_gss14:\*  
119: gb\_gss15:\*  
120: gb\_gss16:\*  
121: gb\_gss17:\*  
122: gb\_gss18:\*  
123: gb\_gss19:\*  
124: cm\_gss13:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.4	11.7	647	40	AI488782
2	191.6	11.5	536	44	AI771830
3	163.6	9.8	649	72	AW442098
4	154.6	9.3	685	63	AW034683
5	137	8.2	546	45	AI897821
6	134.6	8.1	559	45	AI896013
7	134.4	8.0	478	45	AI899102
8	130.4	7.8	500	40	AI489014
9	130.4	7.8	520	40	AI487571
10	128.8	7.7	476	40	AI483541
11	123.6	7.4	747	71	AW349414
12	123	7.4	517	44	AI779099
13	121.4	7.3	658	79	AW651280
14	111	6.6	287	64	AW127679
15	110.8	6.6	537	79	AW650189
16	109.2	6.5	435	79	AW650188
17	107.2	6.4	566	45	AI896026
18	106.6	6.4	518	63	AW032936
19	106.6	6.4	534	72	AW442216
20	106.6	6.4	558	63	AW030266
21	106.6	6.4	576	63	AW031286
22	106.6	6.4	591	79	AW648641
23	106.6	6.4	596	69	AW222352
24	106.6	6.4	605	69	AW220656
25	106.6	6.4	606	63	AW035724
26	106.6	6.4	612	63	AW034633
27	106.6	6.4	613	63	AW035364
28	106.6	6.4	642	79	AW651250
29	106.6	6.4	644	69	AW220874
30	106.2	6.4	496	62	AW439475
31	105	6.3	579	63	AW032414
32	105	6.3	602	63	AW031351
33	104.8	6.3	630	104	AO580287
34	104.6	6.3	627	71	AW398421
35	104.4	6.2	565	63	AW035896
36	103.6	6.2	515	117	AO917129
37	103	6.2	518	63	AW030636
38	102.8	6.2	435	64	AW133020
39	102.2	6.1	466	62	AW440325
40	99.6	6.0	467	45	AI896353
41	98.6	5.9	650	63	AW030131
42	98.4	5.9	534	63	AW031171
43	98.4	5.9	540	44	AI775659
44	97	5.8	543	69	AW216808
45	95	5.7	492	40	AI487283

# ALIGNMENTS

RESULT	1
AI488782	
LOCUS	AI488782
DEFINITION	EST247121 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
	EST 29-JUN-1999
	AI488782 647 bp mRNA
	EST247121 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
	CLED18D16, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AI488782  
AI488782.1 GI:4384153  
EST.  
tomato.  
Lycopersicon esculentum

REFERENCE  
AUTHORS

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 647)  
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vlisson,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.  
Generation of ESTs from tomato carpal tissue  
Unpublished (1999)

TITLE  
JOURNAL  
COMMENT

On Feb 18, 1999 this sequence version replaced gi:4297434.  
Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU.

FEATURES  
source

Location/Qualifiers  
1..647  
/organism="Lycopersicon esculentum"  
/cultivar="TA96"  
/db\_xref="taxon:4081"  
/clone="CLED18D16"  
/clone\_lib="tomato ovary, TAMU"  
/tissue\_type="carpel"  
/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"  
/lab\_host="XLI-Blue MRF"  
/notes="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2: XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."  
BASE COUNT 224 a 97 c 141 g 185 t

Query Match  
Best Local Similarity  
Matches 392; Conservative 0; Mismatches 201; Indels 36; Gaps 4;

11.7%; Score 195.4; DB 40; Length 647;  
62.3%; Pred. No. 3.4e-39;

Qy	632	aatgaagaagaaagtcgacacgctcgacaaagacccgaaagcgaggtctctgataaacac	691
Db	55	ATTCAAAGAGCACTAGACACATTAGATGGTGAAGAAATCCAAAGGTACTTGTGAATAC	114
Qy	692	gttcgacgcatgaagacccgacaaactcaagccatcgacaggtatgaactaatccat	751
Db	115	ATTTGATGCATTAGCTAGAGTACGCCACCTCAAAGCTATTGAAATAATACAAATTTAATTTGGAAT	174
Qy	752	cggccattatccatccatcgattctcagatggcaacgacccctcatcaagcaacaa	811
Db	175	TGGACCATTTGATCTCTCATC---ATTCTGGGTGGAAAGA-----TTTCATTTGGAATC	225
Qy	812	atctacggtggagacacctcttcagaaagccgatgaactacatgactgactgactc	871
Db	226	TTCAATTTGGTGGTGTATCTTTTCAAAAGTCAATGATGACTACATGGAATGGTTAAACAC	285
Qy	872	aaaaccggaatcgcgtctgttacgtttcgttcggagacctcctcgtgagcccgaaacc	931
Db	286	AAAGCCTAAATCATCAATTTGTTATATCTCATTTGGAGTCTATTGTAATTTATCAAGAAA	345
Qy	932	ccaaatgaagaataagcaggtcttcagacacaaatcgccagttctctctggtgat	991
Db	346	CCAAAGAGGAGATTGCCAAAGGTTGATAGATCCAAAGGCCATCTTATGGGTAAAT	405
Qy	992	aagaagaacgaagagggcgacgacaaagcaagcaagcagaagaagaagagctgctgag	1051
Db	406	AAGA-----GATCAAGAAGAAAGAAAGAGAGAGAAATTAAGTTG	447





```

/clone_lib="tomato fruit red ripe, TAMU"
/tissue_type="pericarp"
/dev_stage="red ripe (7-20 days post-breaker)"
/notes="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni; Fruit were tagged at the
breaker stage (first sign of lycopene accumulation on the
blossom end of the fruit) and harvested 7 days
post-breaker (fully red-ripe). 10 days post breaker, and
20 days post-breaker (over-ripe). 20 day fruit which
showed external or internal signs of pathogenesis were
discarded. Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT      217 a 118 c 124 g 190 t
ORIGIN

```

```

Query Match      9.8%; Score 163.6; DB 72; Length 649;
Best Local Similarity 59.3%; Pred. No. 3.8e-31;
Matches 390; Conservative 0; Mismatches 229; Indels 39; Gaps 5;

QY 381 gtgagcttggtggtactctcatctatttttcgtgggcagctgaagtcgcggaagtc 440
Dbb 1 gtcacgctccctcgctacacgtcttctgctacgttggcgagctgaggtagcgcggaactc 60

QY 441 gacgtgcgcagtgccctcttttgattgaacgcgcgtacggttttcgattgtactatttt 500
Dbb 61 CATATCCCATCCGGTACTATGGATTCAACACGCAACTGTGCTAGACATATACTACTAT 120

QY 501 tacttcaatgggtatgcgatgatogta-----tgcgggctcagatca-----a 545
Dbb 121 TACTTCAATGGCTATGAGGATGAATGAAGTGTAGTAGCTCAAAATGATCCAAATGGAGT 180

QY 546 atcaaacgtcccaatctccgcagctctcccaagcaagatctccctcttttctactccct 605
Dbb 181 ATCCAATGCCAGGCTCCATTAATAAAGCCAAAGATCTTCCATCTTTTTAGTTTCA 240

QY 606 tcgagccc-----cgcgagattccgaacccctcatcaagcaaaatgaagaaagtcgac 650
Dbb 241 TCTAGCTCAAAAGATGATAAGTATAGTTTTCCTCTACCAACATTCANAGACCACTAGAC 300

QY 651 acgctcgacaagaacgcgaagcgaaggtcttgatataacacgcttcgcgaattagaacc 710
Dbb 301 ACATTAGATGTTGAAGAAATCCAAAGGTACTTGTGAATACATTTGATGCTATTAGAGTA 360

QY 711 gaacaactcaaacgcacgcagtgatgaactaatatccatcgccattatcccatca 770
Dbb 361 GAGCCACTCAAGCTATTGAATAATACAAATTTAATTTGAATTTGACCATTTGATTCCTTCA 420

QY 771 tcgatattctcagatggcaacgacccctcatcaagcaaaatcctacggtggagacctc 830
Dbb 421 TC---ATTCTTTGGTGGGAAAGAT-----TCATTGGAATCTTCATTTGGTGGTATCTT 471

QY 831 ttcgaaaaacgcgatgaacttacatgagctggcttaaaactcaaaccccgatcatcggtc 890
Dbb 472 TTTCAAAAGTCAAAATGATGACTACATGGAATGGTTTAAACACAAAGCCCTAAATCATCAAT 531

QY 891 gtttacgttttcgagagcctcctcaggtcccgaaaccccaaatggaagaataagca 950
Dbb 532 GTTTATATCTCATTTGGGAGTCTATTGAAATTTATCAAGAAACCAAAAGGAGGAGATTGCA 591

QY 951 atagggtcttcagacaccacacacgcttctctggttgataagaagaacgaagagg 1008
Dbb 592 AAAGGTTGATAGATCCAAAGCCATCTTATGGGTAAATAGAGATCAAGAGAAG 649

```

```

RESULT      4
AW034683    685 bp mRNA EST 15-SEP-1999
LOCUS      EST278414 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION CLECLID5 similar to UDP-glucose:anthocyanin
5-O-glucosyltransferase, putative, mRNA sequence.
ACCESSION  AW034683
VERSION     AW034683.1 GI:5893439

```

```

KEYWORDS      EST.
SOURCE        tomato.
ORGANISM      Lycopersicon esculentum
              Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
              Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
              Solanaceae; Solanum; Lycopersicon.
REFERENCE     1 (bases 1 to 685)
AUTHORS      Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Vision, T.,
              Holt, I.E., Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ahn, S.,
              Ronning, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and
              Giovannoni, J.
              Generation of ESTs from tomato callus tissue
              Unpublished (1999)
              On Apr 3, 1998 this sequence version replaced gi:3018868.
              Contact: David Frisch
              Clemson University Genomics Institute
              Clemson University
              100 Jordan Hall, Clemson, SC 29634, USA
              Tel: 864 656 4366
              Fax: 864 656 4293
              Email: dfrisch@CLEMSON.EDU
              5 prime sequence.
FEATURES      source
              1..685
              /organism="Lycopersicon esculentum"
              /cultivar="TA496"
              /db_xref="taxon:4081"
              /clone="CLECLID5"
              /clone_lib="tomato callus, TAMU"
              /tissue_type="callus"
              /dev_stage="25-40 days old"
              /lab_host="X11-Blue MRF"
              /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
              XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
              of seedlings 7-10 days post-germination were excised, cut
              at both ends and placed on MS medium with no selection.
              Mixed callus was harvested at 25 and 40 days and included
              undifferentiated masses. Tomato Callus EST Library"
BASE COUNT    219 a 128 c 137 g 201 t
ORIGIN

```

```

Query Match      9.3%; Score 154.6; DB 63; Length 685;
Best Local Similarity 59.0%; Pred. No. 7.4e-29;
Matches 378; Conservative 0; Mismatches 224; Indels 39; Gaps 5;

QY 377 tcgctgacgtcttggtggtactctatctatttttcgctggcagctgaagtcgctga 436
Dbb 54 TCCTGTACGTCCTCGTCTACACTCTTCTGTACTTTGGCAGCTGAGGTAGCGGTGA 113

QY 437 agtcgacgtgcgagtgccctcttttgattgaacggcgtacggttttcgattgtacta 496
Dbb 114 ACTCCATATCCATCCCGCTTACTATGATTCAACCCAGCACTGTGCTAGACATATACTA 173

QY 497 ttttactcaatgggtatgcgatgatcgta-----tgcgggctcagatca----- 544
Dbb 174 CTATTACTTCAATGGCTATGAGGATGAATGAAGTGTAGTACTCAATGATGCCAAATG 233

QY 545 ---aatccaactgcccaatcttcgcgagctctccaagcaagatctccctcttttctact 601
Dbb 234 GAGTATCCAAATTTGCCAAGCTTCCATTACTAAAGCCAAAGATCTTCCATCTTTTTAGT 293

QY 602 ccttctgagccc-----cgcgagattccgaacccctcaatgaagaagaagtt 646
Dbb 294 TTCATCTTAGCTCAAAAGATGATAGTATAGTTTTCCTTACCAACATTTCAAAGAGCAACT 353

QY 647 cgacacgtctcgacaaaagaaacgcgaaggtctctgtataaacacggttcgcgcgactaga 706
Dbb 354 AGACACATTAGATGGTGAAGAAATCCAAAGGTACTTGTGAATAGATTGATGCAATTAGA 413

QY 707 aacggaaacactcaaacgcatcgacaggtatgaactaatatccatcgcccaatattcc 766
Dbb 414 GCTAGCGCACTCAAAAGCTATTGAAAAATACAAATTTAATTTGAATTGGACCATTTGATTC 473

```

```

QY 767 atcagatattcagatggcagcaccctcatcaagcaacaaatccctacggtgaga 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 TTATC---ATTCTGGGTGGAAGAT-----TCATTGGAATCTTCATTGGTGTGA 524

QY 827 cctcttcagaaagccgatgaacttacatggactgctaaactcaaaaccccaatcatc 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 TCTTTTTCAAAAGTCAAAATGATGACTACATGGAATGCTTAAACACAAAGCTTAATCATC 584

QY 887 ggtcgtttaagtttcgttcggagcctctgaggtctcccgaaaccccaaatggaagaat 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 585 AATTGTTTATATCTCATTTGGAGTCTATTGCAATTTTCAAGAAACCAAAAGGAGGAGAT 644

QY 947 agcaatagggtcttcagacacaaatcgccagttctctg 987
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 645 TGCAAAAGGGTGTATAGATCCAAAGGCCATCTTATGGG 685

RESULT 5
AI897621
LOCUS EST 546 bp mRNA EST 27-JUL-1999
DEFINITION EST267064 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
CLEDED28K21, mRNA sequence.
ACCESSION AI897621
VERSION AI897621
KEYWORDS EST.
SOURCE GI:5603523
ORGANISM tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 546)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D., and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
On Apr 7, 1998 this sequence version replaced gi:3035329.
TITLE Contact: David Frisch
JOURNAL Lycopersicon esculentum
COMMENT Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.

FEATURES
source Location/Qualifiers
1..546
/organism="Lycopersicon esculentum"
/cultivar="TA96"
/db_xref="taxon:4081"
/clone="CLEDED28K21"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."

BASE COUNT 196 a 82 c 114 g 154 t
ORIGIN

Query Match 8.2%; Score 137; DB 45; Length 546;
Best Local Similarity 58.7%; Pred. No. 2e-24;
Matches 314; Conservative 0; Mismatches 185; Indels 36; Gaps 3;

QY 632 aatgaagaagaagttcagacgctcgacaagaaccgaagcgaggtcttgataaacac 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 ATTCAAAGAGCAACTAGACACATTAGATGGTGAAGAAATCCAAAGGTTACTTGTGATAC 107

```

```

QY 692 gttagcgcattagaaacccgaacaaactcaagccatcgacaggtatgaactaatccat 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 ATTTGATGCTAGAGGTAGAGCCACTCAAGAGCTATTGAAAAATACAAATTTAATGGAAT 167

QY 752 cggcccaatccatccatcgatattcttcagatggcaagcaccctcatcaagaacaa 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168 TGGACCAATTGATTCCTTCATC---ATTCTTGGGTGGAAGA-----TTCATTGGAATC 218

QY 812 atcctacggtggagacctcttcagaaaaagccgatgaacttacatggactgagctaaactc 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 TTCAATTTGGTGGTATCTTTTTCAAAAGTCAAAATGATGACTACATGGAATGGTTAAACAC 278

QY 872 aaaaaccgaatcatcggtcttctggttcggttcgagcctcctgaggtcccgaaacc 931
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 AAAGCCCTAAATCATCAATTCCTTATATCTCATTTGGGAGTCTATTGAATTTATCAAGAA 338

QY 932 ccaaatggaagaaatagcaatagggtcttcagacacacaaatcgccagttctctgggtgat 991
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 CCAAAAGGAGGAGATTGCAAAAGGGTTGATAGAGATCCAAAGGCCATCTTTATGGGTAAAT 398

QY 992 aagaagaaacgaagagggcgacgaacaaagacgaagcgaagaagaagagctgctgag 1051
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 AAGAGATCACGAGAGAAAAGAAAGAGAGAAA----- 432

QY 1052 ctctcttgatcgtcacggaactgaacgactcggaacaaatcgatggtgctcacaatt 1111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 -TTCAGTTGCATGATGGAATTTAGAGAGCAAGGGAATAATAGTACCTGGTGTTCACAAC 491

QY 1112 ggaatgtctgacgcataagtcggtggatgctctgacgcattgcggttggaat 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 492 TGAAGTCTGTACACATCCATCTTTTAGGATGTCTTTGTCTCGCACTGTGGATGGAAT 546

RESULT 6
AI896013
LOCUS EST265456 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION CLEC13H13, mRNA sequence.
ACCESSION AI896013
VERSION AI896013.1 GI:5601915
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 559)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Aho,S.,
Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.

FEATURES
source Location/Qualifiers
1..559
/organism="Lycopersicon esculentum"
/cultivar="TA96"
/db_xref="taxon:4081"
/clone="CLEC13H13"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

```

Xhol: supplier: Giovannoni laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses, Tomato Callus EST Library"

BASE COUNT	201 a
ORIGIN	

**FEATURES**  
**source**

Query Match	8.1%	Score 134.6	DB 45	Length 559
Best Local Similarity	60.0%	Pred. No. 8.3e-24		
Matches 315	Conservative 0	Mismatches 174	Indels 36	Gaps 4
Qy 632	aatgaagaagaagtgcacacgctgcgacaagaagaccgaaagcgaggctctgataaacac	691		
Db 71	ATTCAAAGAGCAACTAGACACATTTAGATGGTGAAGAAATCAAAGACTTGTGGAATAC	130		
Qy 692	gttcgacgcattagaaccgaacactcaagccatcgacaggtatgaactaatccat	751		
Db 131	ATTTGATGCATTAGCTAGAGCCACCTCAAGCTATTGAAAATAACAATTTAATTGGAAAT	190		
Qy 752	cggcccattaatccatcatcgatattctcagatggcaacgccctcatcaagcaaca	811		
Db 191	TGGACCATTGATCTCTTCATC--ATTCCTGGGTGGAAGAA-----TTCAATGGGAATC	241		
Qy 812	atctacggtggagacctcttcagaaaaagcgcgatgaacctacatggactggctaaacc	871		
Db 242	TTCAATTTGGTGTGATCTTTTTCAAAGTCAAATGATGACTACATGGAATGGTTAAACAC	301		
Qy 872	aaaccgcgaatcatcgctgctttacgttcttcgaggacctcctcgaggtccccgaacc	931		
Db 302	AAAGCCTAAATCATCAATTTGTTTATCTCATTTGGGAGTCTATTGAATTTATCAAGAAA	361		
Qy 932	ccaaatggagaatatgacaataggcttccagacaccaaactcgccagctctctgggtgat	991		
Db 362	CCAAAGGAGGAGGATTGCCAAAGGGTTGATAGATCCAAAGGCCATCTCTTATGGGTAAAT	421		
Qy 992	aagaagaacgaagggcgacgcacaagagcgaagcagaagaagaagagagctgctgag	1051		
Db 422	AAGA-----GATCAAGAAGAGAAAAGAAAGAGAGAATAAAGTTG	463		
Qy 1052	ctctcttgatcgtcacggaactgaacgactcgggaaaaactcgtgacatggtgctcacaa	1111		
Db 464	CATGATGGA-----ATTAGAGAACGAGGGAATACTACCATCGTGTTCACAACT	514		
Qy 1112	ggaagtcctgacgcataagtcggtgggatgcttcgtgacgcattg	1156		
Db 515	TGAAGTCTCAGACATGCATCTTTAGATGATGTTGCTCTCGCACTG	559		

RESULT	7
AI899102	
LOCUS	478 bp mRNA EST 27-JUL-1999
DEFINITION	EST268545 tomato ovary, TAMU Lycopersicon esculentum cDNA clone CLED37A16, mRNA sequence.
ACCESSION	AI899102
VERSION	AI899102.1 GI:5605004
KEYWORDS	EST.
SOURCE	tomato.
ORGANISM	Lycopersicon esculentum Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE	1 (bases 1 to 478)
AUTHORS	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujiki,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J. Generation of ESTs from tomato carpel tissue. Unpublished (1999).
TITLE	
JOURNAL	
COMMENT	On Feb 18, 1999 this sequence version replaced gi:4296853. Contact: David Frisch Clemson University Genomics Institute

```

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
Location/Qualifiers
1. 478
/organism="Lycopersicon esculentum"
/cultivar="TA498"
/db_xref="taxon:4081"
/clone="cLED37A16"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="X11-Blue MR#"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
172 a 67 c 137 t
BASE COUNT

```

RESULT	8
AI489014	
LOCUS	EST
DEFINITION	29-JUN-1999
	AI489014 500 bp mRNA
	EST247353 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
	CLED19F7, mRNA sequence.
ACCESSION	AI489014
VERSION	AI489014.1 GI:4384385







On Jan 6, 2000 this sequence version replaced gi:6677182.  
Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU

```

FEATURES             Location/Qualifiers
     source            1..658
                        /organism="Lycopersicon esculentum"
                        /cultivar="TA496"
                        /db_xref="taxon:4081"
                        /clone="GLEI16K15"
                        /tissue_type="whole seedlings"
                        /dev_stage="7 days post imbibition"
                        /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; 7 days post imbibition on water-agar. Mixed stage whole germinating seedlings from seed coat emergence up to two centimeters in seeds not showing obvious signs of germination were discarded."
BASE COUNT           208 a   126 c   124 g   200 t
ORIGIN
Query Match          7.3%; Score 121.4; DB 79; Length 658;
Best Local Similarity 57.9%; Pred.No.1.9e-20;
Matches 332; Conservative 0; Mismatches 201; Indels 40; Gaps

Qy 377 tcgcgtgagctttgggtgactctcatctatttcgtggcgacgtgaagtgccgctga 436
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88 TCCTGTACGTCCTCGCTACACTCTTCTGCTACCTTGGGCAGCTGAGGTAGCGGTGA 147

Qy 437 agtcagcgtccgagtgcccttcttggattgaaccggtcacggttttcgatgtgta 496
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 ACTCCATATCCCATCCGCCTTACTGAATTCAACCAGCAACTGTGCTAGACATATACTA 207

Qy 497 ttittactcaatgggtatgccgatgatcgaa---tgcggggtcagatca----- 544
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 208 CTATTACTTCAAATGGGTCTAGAGTAGAAATGAAGTAGTACTACCTCAAATGCCAAATTG 267

Qy 545 ---aatccaactgcccaatcttcgcagctctccaagcaagatctcccctcttccact 601
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 GAGTATCCAATTGCCAAGGCTTCCATTACTAAAAGCCAAGATCTTTCCATCTTTTTTAG 327

Qy 602 cccttcgagccccgcgagat-----tcggaaccttaataagaagaagt 645
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 328 TTTTCATCTAGCTCAAAAGATGATGAGTATAGTTTGTCTCTACCAACATTCAAAGAGCAAC 387

Qy 646 tcgacacgctcgacaagaagaccgaagcgaaggtcttgataaacaggttcgcagcattag 705
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 TAGACACATTAGATGTTGGAAGAAAAATCCAAAGGACTTGTGAATACATTGATGGCATAG 447

Qy 706 aaaccgacaaactcaaagccatcgacaggtatgaactaatccatcgccccaattaacc 765
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 448 AGCTAGAGCCACTCAAGCTATTGAAAAATACAATTAATTGGAATTGACCATTGATTC 507

Qy 766 catcatcgatatctcagatggcacaagccctcatcaagcaacaatacctcagcgtggag 825
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 508 CTTTCATC--ATTCTTGGGTGGAAGA-----TTTCATTGGAATCTTCATTGGTGGGTG 558

Qy 826 acctcttcagaagaagccgatgaacttacatggactggctaactcaaacccgaatacat 885
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 559 ATCTTTTTCAAAAGTCAAAATGATGACTACATGGAAATGGTTAAACACAAAAGCCTAAATCAT 618

Qy 886 cgfgcgtttacgtttcgttcgggagcctctga 918
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 619 CAATTGTTTATATCTCATTAGGAGTCTATTGA 651

```

AW127679	287 bp	mRNA	EST	25-OCT-1999
LOCUS				
DEFINITION	M110425 DSLC Medicago truncatula cDNA clone L244,			mRNA sequence.
ACCESSION	AW127679			
VERSION	AW127679.1	GI:6115583		
KEYWORDS	EST.			
SOURCE	barrel medic.			
ORGANISM	Medicago truncatula			
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.			
AUTHORS	1 (bases 1 to 287)			
TITLE	Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Gonzales,M.B. and Ellis,L.			
JOURNAL	ESTs from leaf/cotyledon cDNA library of Medicago truncatula			
COMMENT	Unpublished (1999)			
	On Jul 19, 1995 this sequence version replaced gi:2939533.			
	Contact: Carroll P. Vance			
	Department of Agronomy and Plant Genetics			
	University of Minnesota			
	411 Hortlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA			
	Tel: 612-625-7219			
	Fax: 651-649-5058			
	Email: vance004@maroon.tc.umn.edu			
	Developmental stage: mature leaves and young cotyledon; Date: 9/15/99; Submitted to the Database of Expressed Sequence Tags (dbEST) on 10/22/99; More information is available at 'http://chryslie.tamu.edu/medicago'.			
FEATURES	Seq primer: T3.			
source	Location/Qualifiers			
	1..287			
	/organism="Medicago truncatula"			
	/cultivar="genotype A17"			
	/db_xref="taxon:3880"			
	/clone="L244"			
	/clone_lib="BSLC"			
	/tissue_type="leaf + cotyledon"			
	/note="Vector: pBluescript SK +/-; Site 1: EcoRI; Site 2: XhoI; one month old N2 fixing root nodules"			
BASE COUNT	78 a 29 c 96 g 84 t			
ORIGIN				
	Query Match 6.6%; Score 111; DB 64; Length 287;			
	Best Local Similarity 63.9%; Pred. No. 7 le-18;			
	Matches 168; Conservative 0; Mismatches 95; Indels 0; Gaps 0;			
QY	1035 gaagagaagctgtcgtgactcttcttgatcgtcacgaactgaacgactcgggaaatcgtg 1094			
DB	12 GATGATGATGAGCTGAGTCTGTAGAGAGGAACCTTGAGAATAATATGAATGGGAAGATAGTA 71			
QY	1095 acatggtgcacaaatcggatgtctcagcataaagtcggtgggatactctcgtgaacgcat 1154			
DB	72 AAATGGTGTAGTCAGGTGGAGGTTTTCGACACCGTTCTTTGGGTTGTTTATGACCAT 131			
QY	1155 tcggttggaattcgtatcagagacgtggctgtggtggtgcgcgtggtggtgcttct 1214			
DB	132 TGTGGATGGAATTCGACTTTGGAAGGTTTGGGTCACGGGTTCTCTATGTTGCATTTCT 191			
QY	1215 caatggttcgatacaaggagcgaatcgaagaatgatcgaagaatgctgtaggagtggtg 1274			
DB	192 CAGTGGACAGATCAACCACTAATGCCAAGTTAATTGAAGATGTGTGGAAGACTGGGTTG 251			
QY	1275 agagtcagatgaatgagggaag 1297			
DB	252 AGGATGGAACATGATGAGGAAG 274			
RESULT 15				
AW50189/c				
LOCUS	AW50189	537 bp	mRNA	EST
DEFINITION	EST328643 tomato germinating seedlings, TAMU Lycopersicon esculentum cDNA clone cLPI112c8 5', mRNA sequence.			04-APR-2000



Job time: 62707 sec

THE UNIVERSITY OF CHICAGO

Search completed: August 1, 2000, 14:38:35

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:50:22 ; Search time 47.63 Seconds  
(without alignments)  
237.706 Million cell updates/sec

Title: US-09-147-955-8  
Perfect score: 2484  
Sequence: 1 MVNKRHILLATFPAQGHNP.....GSSMNLKNEITRIENAS 478

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues  
Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2480	99.8	478	1 W92950	WO9905287 Seq ID 1
2	1473.5	59.3	461	1 W92949	WO9905287 Seq ID 9
3	1445	58.2	460	1 W92947	WO9905287 Seq ID 7
4	1412.5	56.9	443	1 W92948	WO9905287 Seq ID 8
5	1178.5	47.4	468	1 W92952	WO9905287 Seq ID 1
6	631	25.4	466	1 W92925	UDP-glucose:thiohy
7	631	25.4	471	1 W03756	Maize UDP-glucose
8	420	16.9	433	1 W13406	Solanum melongena
9	393.5	15.8	471	1 W98009	Medium-chain UDP-g
10	392.5	15.8	470	1 W47172	Glucosyl transfera
11	382	15.4	488	1 W64569	Potato solanine
12	351.5	14.2	473	1 W58451	UDP-glucose:flavon
13	322	13.0	471	1 R49245	Anthocyanidin-3-gl
14	164.5	6.6	534	1 R26154	HUG-Br2. Isolated
15	156	6.3	533	1 R26153	HUG-Br1. Isolated
16	148.5	6.0	506	1 R10429	Ecdysteroid UDP-gl
17	145	5.8	515	1 W58750	Ecdysteroid UDP-gl
18	131.5	5.3	530	1 W47126	Uridine diphospho-
19	109.5	4.4	489	1 W20344	H. pylori putative
20	106.5	4.3	957	1 W20841	H. pylori transmem
21	99.5	4.0	617	1 R15241	B. lautus endol cor
22	98	3.9	473	1 W64392	S. cerevisiae L947
23	98	3.9	1198	1 W64384	S. cerevisiae L947
24	97.5	3.9	339	1 Y07063	Renal cancer assoc
25	97.5	3.9	1006	1 W77290	Human differentiat
26	96.5	3.9	700	1 R13227	Novel endoglycanas
27	96	3.9	906	1 R42354	Human glutamate re
28	96	3.9	906	1 R84913	Human glutamate re
29	96	3.9	907	1 R11989	Glutamate receptor
30	94.5	3.8	414	1 R39982	Saccharothrix aero
31	94.5	3.8	865	1 R20670	Lipoxygenase. Prod
32	94.5	3.8	865	1 R23797	Rice lipoxygenase.
33	91.5	3.7	74	1 R30165	UGT1 Exon 4 produc

34	90.5	3.6	1162	1 W19115	Murine long form O
35	89.5	3.6	574	1 Y00940	Aquifex VF-5 DNA p
36	89	3.6	2161	1 W87724	Rupestis stem plc
37	89	3.6	2710	1 R95016	C. difficile toxin
38	89	3.6	2710	1 W68387	Clostridium diffie
39	88.5	3.6	1015	1 W34259	Rat ob receptor is
40	88.5	3.6	1162	1 W23399	Rat ob receptor (w
41	88.5	3.6	1162	1 W23398	Rat ob receptor (w
42	88.5	3.6	1162	1 W34257	Rat wild-type ob r
43	88.5	3.6	1284	1 R78187	Sequence encoded b
44	88	3.5	409	1 R78525	Protein A-calmodul
45	87	3.5	1151	1 W77287	zebrafish differen

ALIGNMENTS

RESULT 1

W92950  
ID W92950 standard; Protein; 478 AA.  
AC W92950:  
DT 14-MAY-1999 (first entry)  
DE WO9905287 Seq ID 10.  
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.  
OS Torenia hybrira.  
FH Key  
FT Misc\_difference 64 Location/Qualifiers  
FT Misc\_difference 65 /note= "Cys or Phe"  
FT Misc\_difference 65 /note= "Ser or Pro"  
PN WO9905287-Al.  
PD 04-FEB-1999.  
PF 16-JUL-1998; J03199.  
PR 25-JUL-1997; JP-200571.  
PA (SUNR ) SUNTORY LTD.  
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
PI Yamazaki M;  
DR WPI: 99-142940/12.  
DR N-PSDB; X02829.  
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration  
PS Claim 2; Page 78-81; 89pp; Japanese.  
CC This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut flowers harvested from them.  
CC Sequence 478 AA;  
SQ

Query Match 99.8%; Score 2480; DB 1; Length 478;  
Best Local Similarity 100.0%; Pred. No. 6.3e-236;  
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVNKRHILLATFPAQGHNP	SLFAKRLNTGYVDQVTFSTVYALRRMFE	DPSSRID	60
Db	1	MVNKRHILLATFPAQGHNP	SLFAKRLNTGYVDQVTFSTVYALRRMFE	DPSSRID	60
Qy	61	FVAXXSDYDGLKKGGDK	KNYMSEMRKGT	KALKDTLKLNDAA	MGSECYNRVSVYVSH
Db	61	FVAXXSDYDGLKKGGDK	KNYMSEMRKGT	KALKDTLKLNDAA	MGSECYNRVSVYVSH
Qy	121	LFSWAAEAREVDVPSALL	WIEPATVDFVYFNGVADD	IDAGSDQIQLPNL	PQLSKQD
Db	121	LFSWAAEAREVDVPSALL	WIEPATVDFVYFNGVADD	IDAGSDQIQLPNL	PQLSKQD
Qy	181	LPSEFLPSSPARETLMKE	FDTLDPKAKVLINFDAL	ETQLKADRYELISG	PGLI
Db	181	LPSEFLPSSPARETLMKE	FDTLDPKAKVLINFDAL	ETQLKADRYELISG	PGLI
Qy	241	PSIFSDGNDPSSSNKSY	GGDLFRKADETYMDWLN	SKPSSVSVVFGSL	LRPKPQME
Db	241	PSIFSDGNDPSSSNKSY	GGDLFRKADETYMDWLN	SKPSSVSVVFGSL	LRPKPQME

QY 301 IAIGLSDTKSPVLWVIRRNEDGDEQAEAEKLLSPFDRHGTGRGLKIVTWCSDQDVLIT 360  
DB 301 IAIGLSDTKSPVLWVIRRNEDGDEQAEAEKLLSPFDRHGTGRGLKIVTWCSDQDVLIT 360  
QY 361 HKSVGCVFTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKIMIEDVWRSVGRVVRNEEGV 420  
DB 361 HKSVGCVFTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKIMIEDVWRSVGRVVRNEEGV 420  
QY 421 VDRREIKRCVSEVTKSRELRESAMWGLAKEAMDEERGSSMNLKNFTRIIENAS 478  
DB 421 VDRREIKRCVSEVTKSRELRESAMWGLAKEAMDEERGSSMNLKNFTRIIENAS 478  
RESULT 2  
W92949  
ID W92949 standard; Protein; 461 AA.  
AC W92949;  
DT 14-MAY-1999 (first entry)  
DE W0905287 Seq ID 9.  
KW Plant: flavonoid 5-transglycosylation activity; 5TG; variety: colour.  
OS Verbena hybrida.  
PN W0905287-A1.  
PD 04-FEB-1999.  
PF 16-JUL-1998; J03199.  
PR 25-JUL-1997; JP-200571.  
PA (SUNR ) SUNTORY LTD.  
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
PI Yamazaki M;  
DR WPI; 99-142940/12.  
DR N-PSDB; X02828.  
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
PT used to transform plants for improvement of plant coloration  
PS Claim 2; Page 76-78; 89pp; Japanese.  
CC This invention describes the isolation of plant proteins which have  
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used  
CC in the production of plant varieties with improved colour, for example in  
CC the production of flower varieties with desired coloration and cut  
CC flowers harvested from them.  
SQ Sequence 461 AA;

Query Match 59.3%; Score 1473.5; DB 1; Length 461;  
Best Local Similarity 61.5%; Pred. No. 8.8e-137;  
Matches 295; Conservative 71; Mismatches 85; Indels 29; Gaps 11;  
QY 2 VNKHILLATPPAGCHINPSLEPAKRLNTGYVDQVTFSTVYALRRM-RETPDSSRID 60  
DB 1 MSRAHVLLATPPAGCHINPALQFAKRLNADI--QVTFSTVYARRRMSRTAAGSGLIN 58  
QY 61 FVAXXDSYDDGLKGGDKNVMSEMRKGTALKDITLIKLNDAAMGSECYNRVSVFYVSH 120  
DB 59 FVSFSDYDDGLQDGGDKNVMSEMRKGTALKSDTLAANNVDKSS---KITFVYVSH 114  
QY 121 LFSWAAEVAEVDVPSALLMIEPATVDFVYFNGYADDIDAGSDQIQLP-NLPQLSKQ 179  
DB 115 LFAWAAKVAEFLHRSALLMIEPATVLDIPYFNGYSDEIDAGSDAHLPGGLPVLQAR 174  
QY 180 DLPSEFLPSPAPRTLMKEFDTLDEKPAKVLINTFDALTEQLKAIDRYELISIGPL 239  
DB 175 DLPSEFLPSPAPRTLMKEFDTLDEKPAKVLINTFDALTEQLKAIDRYELISIGPL 234  
QY 240 IPSSIFSDGNDPSSSNKSYGGDLFRKA--DFTYMDLNSKPSSVYVSGSLRLRPKPQ 297  
DB 235 IPSA-FLDGRDP--SDRSFGDGLFEKSNDDCLWELSTNPRSSVYVSGSFYNTTSQ 291  
QY 298 MEETAIIGLSDTKSPVLWVIRRNEDGDEQAEAEKLLSPFDRHGTGRGLKIVTWCSDQD 357  
DB 292 MEETAIIGLSDTKSPVLWVIRRNEDGDEQAEAEKLLSPFDRHGTGRGLKIVTWCSDQD 340  
QY 358 VLTHKSGCVFTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKIMIEDVWRSVGRVVRNEE 417  
DB 341 VLTHKSGCVFTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKIMIEDVWRSVGRVVRNEE 400

QY 418 GGVVDRREIKRCVSEVI---KSRELRESAMWGLAKEAMDEERGSSMNLKNFTRII 473  
DB 401 GSVVDGDEIRKCIIEVMDGGEKSRRLRESACKWDLAKAMEED-GSSVNLKVLDEVV 459  
RESULT 3  
W92947  
ID W92947 standard; Protein; 460 AA.  
AC W92947;  
DT 14-MAY-1999 (first entry)  
DE W0905287 Seq ID 7.  
KW Plant: flavonoid 5-transglycosylation activity; 5TG; variety: colour.  
OS Perilla frutescens.  
PN W0905287-A1.  
PD 04-FEB-1999.  
PF 16-JUL-1998; J03199.  
PR 25-JUL-1997; JP-200571.  
PA (SUNR ) SUNTORY LTD.  
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,  
PI Yamazaki M;  
DR WPI; 99-142940/12.  
DR N-PSDB; X02826.  
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
PT used to transform plants for improvement of plant coloration  
PS Claim 2; Page 71-73; 89pp; Japanese.  
CC This invention describes the isolation of plant proteins which have  
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used  
CC in the production of plant varieties with improved colour, for example in  
CC the production of flower varieties with desired coloration and cut  
CC flowers harvested from them.  
SQ Sequence 460 AA;  
Query Match 58.2%; Score 1445; DB 1; Length 460;  
Best Local Similarity 61.1%; Pred. No. 5.6e-134;  
Matches 292; Conservative 65; Mismatches 89; Indels 32; Gaps 11;  
QY 4 KRHILLATPPAGCHINPSLEPAKRLNTGYVDQVTFSTVYALRRMRETPDSS----RI 59  
DB 3 RRVLLATPPAGCHINPALQFAKRLKAG--TDVTFSTVYARRRMTASAAAGNPGL 60  
QY 60 DFVAXXDSYDDGLKGGDKNVMSEMRKGTALKDITLIKLNDAAMGSECYNRVSVFYVYS 119  
DB 61 DFVAFSDYDDGLKPGCGDKNVMSEMRKGTALKDITLIKLNDAAMGSECYNRVSVFYVYS 110  
QY 120 HLFSAWAAEVAEVDVPSALLMIEPATVDFVYFNGYADDIDAGSDQIQLPNLPQLSKQ 179  
DB 111 HLFSAWAAEVAEVDVPSALLMIEPATVDFVYFNGYADDIDAGSDQIQLPNLPQLSKQ 170  
QY 180 DLPSEFLPSPAPRTLMKEFDTLDEKPAKVLINTFDALTEQLKAIDRYELISIGPL 239  
DB 171 SLPTFLPSPAPRTLMKEFDTLDEKPAKVLINTFDALTEQLKAIDRYELISIGPL 230  
QY 240 IPSSIFSDGNDPSSSNKSYGGDLFRKADE-TYMDLNSKPSSVYVSGSLRLRPKPQ 298  
DB 231 IPSA-FLDGGDPSET--SYGGDLFEKSEENCNVEWLDTPKSSVYVSGSVFLRPKQ 287  
QY 299 BEETAIIGLSDTKSPVLWVIRRNEDGDEQAEAEKLLSPFDRHGTGRGLKIVTWCSDQD 358  
DB 288 BEETAIIGLSDTKSPVLWVIRRNEDGDEQAEAEKLLSPFDRHGTGRGLKIVTWCSDQD 341  
QY 359 LTHKSGCVFTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKIMIEDVWRSVGRVVRNEEG 418  
DB 342 LAHPALGCVFTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKIMIEDVWRSVGRVVRNEEG 401  
QY 419 GVVDRREIKRCVSEVI---KSRELRESAMWGLAKEAMDEERGSSMNLKNFTRII 472  
DB 402 G-VDSIERCVCVENMDGGEKSKLVRENAIKWKTAREMGED-GSSLKLNALFLHQV 457  
RESULT 4  
W92948





(RESE ) RESEARCH CORP TECHNOLOGIES INC.  
 PI Bandurski RS, Szczygłowski K, Szerszen JB;  
 DR WPI: 96-068875/07.  
 DR N-PSDB; T10106.  
 PT UDP-glucose indol-3-yl:acetyl:glucosyl transferase DNA and protein  
 PT - used in sense or anti-sense orientation for the genetic control of  
 PT plant IAA levels and plant growth  
 PS Claim 3; Page 21-23; 41pp; English.  
 CC This sequence encodes a UDP-glucose-indol-3-  
 CC ylacetylglucosyltransferase (IAGlu-transferase) from maize (Zea mays).  
 CC Over-expression of this sequence in transgenic plants will result  
 CC in at least 1 of the following effects: inhibition of apical  
 CC dominance; inhibition of stem elongation; inhibition of cell  
 CC enlargement and increased numbers of stems per plant as compared  
 CC with a wild-type plant. The ability to control the expression of  
 CC this enzyme allows the control of free indoleacetic acid levels in  
 CC plants, thereby effecting plant growth rates. Antisense constructs  
 CC may be used to inhibit synthesis of this enzyme.  
 SQ Sequence 471 AA;

Query Match 25.4%; Score 631; DB 1; Length 471;  
 Best Local Similarity 33.7%; Pred. No. 9.4e-54;  
 Matches 163; Conservative 78; Mismatches 197; Indels 46; Gaps 16;

QY 6 HILLATPPAGCHINPSLEFAKRLNTGCVDOVTFFTSVALRRMRFTDPSRIDFVAXX 65  
 DB 4 HVLVPPFGGHNPNVQFAKRLASKVA--TTLVTRFRTQRTADVDAHAM--VEAIS 58  
 QY 66 DSYDDGLKGGDKGNMSEMRKGTALKDTLIKLNDAAMGSECYNRVSVYVSHLFSWA 125  
 DB 59 DGHDEG---GFASAAGVAEYLEKQAAAAASLASLVEARASSA--DAFTCVVYDSYEDWV 113  
 QY 126 AEVAREVDVPSALLNTEPATVDFVYFYFNG-----YADDIDAGSDIQLP-----NL 173  
 DB 114 LPVARRMGLPAVPFSTQSCAVSAVYHFSGRLAVPPGAAADGSDGAGAAALSEAFGL 173  
 QY 174 POLSKODLPFLPSSPARFRTLMKEKFDPLDKEPRAK-VLINTFOALETEQKADRY - 231  
 DB 174 PEMERSELPSEFVDHGP-YPTIMAQAIQFAHAGKDDWVLFNSFELETEVLAKTKYL 231  
 QY 232 ELISIGLIPSSIFSGNDPSSNNK-SYGGDLFRKADETYMDLNSKPSSVYVYVSGSL 290  
 DB 232 KARAGPCVP-LPTAGRTAGNRTYGANLV-KPEDACTKWLDTKPDRSVAYVSEGL 288  
 QY 291 LRLPKQMEIAIGLSDTKSPVLVIRNREEGDEQAEKEELKLSFFDRHGTER-LGKI 349  
 DB 289 ASLGNQKEELARGLLAAGKPFLLWYVR-----ASDEHQVPRYLLAEATATCAAMV 338  
 QY 350 VTWCSOLDVLTHKSGVCFVTHCGWNSAIESLACGVPVVCPPQWFDQGTNAKMIEDVWRSG 409  
 DB 339 VPMCPQLDLVAHPAVGCFVTHCGWNSLEALSPGVPMVAMALTDQPTNARNVELAWAG 398  
 QY 410 VYRVNREEGVDRREIKRCVSVIKRSRE----LRSEAMWKLAKAMEDEERGSSMNNL 465  
 DB 399 VRARRDAGAGVFLRGVEVCRAVMDGGEAAASAARKAAGWRDRARA-V-APGSSDRNL 457  
 QY 466 KNFI 469  
 DB 458 DEFV 461

RESULT 8  
 W13406  
 ID W13406 standard; Protein; 433 AA.  
 AC W13406;  
 DT 10-JUN-1997 (first entry)  
 DE Solanum melongena flavanoid-3-glucosyl transferase.  
 KW Eggplant; flavanoid-3-glucosyl; transferase; glucose; flavonoid;  
 KW pigment; production; seed; plant.  
 OS Solanum melongena.  
 PN J09056385-A.  
 PD 04-MAR-1997.

PF 25-AUG-1995; 238943.  
 PR 25-AUG-1995; JP-238943.  
 PA (KIRI ) KIRIN BREWERY KK.  
 DR WPI: 97-206631/19.  
 DR N-PSDB; T62123.  
 PT Solanum melongena flavanoid-3-glucosyl transferase and DNA encoding  
 PT it - catalyses transfer of glucose to 3-position of flavonoid  
 PT pigment in plants  
 PS Claim 4; Pages 9-10; 14pp; Japanese.  
 CC The present sequence is the Solanum melongena (eggplant)  
 CC flavanoid-3-glucosyl transferase (F3GT), which catalyses the  
 CC transfer of glucose to the 3-position of flavonoid pigments. The  
 CC F3GT DNA is useful for the production of the seed of a plant. A  
 CC S. melongena seed was cultured under a red light and a flavanoid  
 CC derived. A cDNA library in which the pigment was concentrated was  
 CC prepared. A flavanoid glucosidising enzyme gene was cloned and its  
 CC structure and mRNA expression pattern analysed. The clone was  
 CC expressed in E. coli.  
 SQ Sequence 433 AA;

Query Match 16.9%; Score 420; DB 1; Length 433;  
 Best Local Similarity 27.1%; Pred. No. 5.1e-33;  
 Matches 136; Conservative 93; Mismatches 161; Indels 112; Gaps 26;

QY 6 HILLATPPAGCHINPSLEFAKRLNTGCVDOVTFFTSVALRRMRFTDPS-----SRID 60  
 DB 7 HIAFLAPFCTHATPLTLVQKI--SPFLPSSSTIFSP-----FWTSSNSIFSKVP 56  
 QY 61 FVAXXDSYD--DGLKGGDKGNMSEMRKGTALKK----DTLI--KLNDAAMGSECYNR 112  
 DB 57 NOENIKIYNVMDGVKGEKD-----TPFGLAIAKLFIOSTLLISKITEEA---EETG 105  
 QY 113 VSP-VVYSHLFSWA--AEVAREVDVPSALLNTEPATVDFVYFYFNGYADDIDAGSD-QI 168  
 DB 106 VKFSCIFSDAFLMCFVLKPKKMAPGVAYWTGGSCSLAVHL-----YTDLIRSNKETSL 160  
 QY 169 OLPNL-----POLSKODLPFLPSSPARFRTLMKEKFDPLDKEPRAKVLINTF 217  
 DB 161 KIFGFSSTLSINDIPPEVTAEDLEG---PMSSMLYNALNLHKADA-----VVLNSF 209  
 QY 218 DALETEOL--KAIDR--YELISIGLIPSSIFSGNDPSSNNKSYGGDLFRKADET-YMD 272  
 DB 210 QELDRDPLINKLQKLNQKVFNIGPLVLQS-----SRKLDSESGCIQ 250  
 QY 273 WLNSKPSSVYVYVSGSLRLKLPQMBEIEAIGLSDTKSPVLVIRNREEGDEQAEKEE 332  
 DB 251 WLDKQKEKSVVYLSFGVTTLPPNEIGSIAEALETKTTPFIWSLRNNGVKNLPK----- 304  
 QY 333 KLISFPDRHGTERTLKGIVTWCSQLDLVTHKSGVCFVTHCGWNSAIESLACGVPVVCPPQW 392  
 DB 305 ---GFLER--TKFEGKITVSWAPQLEILAHKSVGVFVTHCGWNSILEGISTGVPMICRPF 359  
 QY 393 FDQGTNAKMIEDVWRSGVYRVNREEGVDREIKRCVSEVI---KSRLERESAMWKL 449  
 DB 360 GDQKLSRPMVESWEIGLQI---EGGIFTKSGIISALDTFFNEEKGLIRENV---EGL 412  
 QY 450 AKEMDEERGSSMNNLNKFNITR 471  
 DB 413 KERAL--EAVNQMWQVQOKISR 432

RESULT 9  
 W98009  
 ID W98009 standard; Protein; 471 AA.  
 AC W98009;  
 DT 21-JUN-1999 (first entry)  
 DE Medium-chain UDP-glucose:fatty acid glucosyltransferase.  
 KW UDP-glucose:fatty acid glucosyltransferase; tomato;  
 KW 1-O-acyl-a-glucose; acyl donor; esterification.  
 OS Lycopersicon pennellii.  
 OS Lycopersicon esculentum.  
 PN W09909144-A1.

PD 25-FEB-1999.  
PF 10-AUG-1998; IBI369.  
PR 30-JUN-1998; US-106464.  
PR 13-AUG-1997; US-055554.  
PA (CORR ) CORNELL RES FOUND INC.  
PI Eannetta NT, Ghangas GS, Kua J, Steffens JC;  
DR WPI; 99-190155/16.  
DR N-PSDB; X24873.  
PT New chain length specific uridine diphosphate-glucose:fatty acid  
PT glucosyltransferases - useful for preparing straight or branched,  
PT (un)saturated 1-o-acyl-a-glucoses and 1-o-acyl-a-D-glucoses  
PS Example 3; Page 36-38; 42pp; English.  
CC The present sequence is a UDP-glucose:fatty acid glucosyltransferase  
CC that transfers glucose to medium-chain length fatty acids to form  
CC 1-o-acyl-a-glucoses. The 1-o-acyl-a-glucoses act as acyl donors  
CC in the esterification of glucose and in the further esterification  
CC of partially acylated glucose in reactions catalysed by glucose  
CC acyltransferases. They also act as acyl donors in the  
CC esterification of sucrose and in the further esterification of  
CC acyltransferases. A cDNA clone (see X24873) encoding the enzyme  
CC was isolated from an FI population of a leaf trichome cDNA library  
CC prepared from an FI population of a cross between wild tomato  
CC Lycopersicon pennellii (LA 716) and cultivated tomato  
CC Lycopersicon esculentum cv. New Yorker. The invention provides  
CC a chain length-specific UDP-glucose:fatty acid glucosyltransferases  
CC and a method for their purification. The enzymes have specificity  
CC for transferring glucose to short, medium or long chain length  
CC fatty acids. Methods for preparing 1-o-acyl-a-D-glucoses, where  
CC the acyl group comprises 3-5, 6-13 or 14-22 C atoms, are claimed.  
SQ Sequence 471 AA;

Query Match 15.8%; Score 393.5; DB 1; Length 471;  
Best Local Similarity 26.2%; Pred. No. 2.4e-30;  
Matches 136; Conservative 78; Mismatches 186; Indels 125; Gaps 23;

QY 6 HILLATPAQGHINPLEFAKRLNTGVVDQVTFET-----SVVA-----L 46  
DB 6 HFFFPMAQGHMPTLDMALVACRGV--KATITPLNESVFSKAIERKHHGIEIDI 63  
QY 47 RRMRFET-----DPSSRIDFVAXXSDYDDGLKGGDKNMYSEMRRKGTALKADTLKL 100  
DB 64 RLLKFPKAKNDLPEDCERLDLPVSDDKLPNFK-----AAAMMKDEPEEL 108  
QY 101 NDAAWGSECVNRVSVVYSHLFSMAAEVAREVDVPSALLWIEPATVDFVYFYP----- 154  
DB 109 I-----GEC---RPDCLVSDMFLPWTTSAAKFSIPRIVF-----HGTSYFALCVGHSI 154  
QY 155 --NGYADDIAGSDIOIOLNPLP---QLSKODLPSPFLPSPAPFRTLMKKEFDTLDKPK 209  
DB 155 RKNPFKNVSDTETVFPDPFHEIRLTRTQLSPFEQSDTEGTMAPHIKAV-----RES 209  
QY 210 AK---VLINTDALETOLKAIDRYELI-----SIGLIPSSIFSDGNDPSSNKK--SY 258  
DB 210 AKSYGVIFNFSYELESY---VEHYTKVVGKKNWAIGPL-----SLCNRDIET 254  
QY 259 GGDLPKRA---DEYMDWLNKSPSSVVYVSGSLRLPKPOMEIATGLSDTKSPVLWV 315  
DB 255 KAERGRKSSIDEHAELKWLDSKSSIVYVCGFSTADFTTAQOMELAMGLEASGQDFI 314  
QY 316 IRRNEEGDEOQAEERKLLSFDRHGTERTLGKIVT--WCSOLDVLTHKSVGCFVTHCGWN 374  
DB 315 IRTGNEWLPFGFEER-----TKEGLIIRGWAPVLIDHBAIGAFVTHCGWN 363  
QY 375 SAIESLACGVVVPVFPQWPDGQTNKMIEDYWRSGVRV-----RVNEEGGVVDRREIKR 428  
DB 364 STLGEISAGVPLMTWPVFAEQEFNEKLVTEVMRSGAGVSGQWKRTASEG--VKREATAK 421  
QY 429 CVSEVIKRSRE---LRRESAMMKGLAKEMDEERSSNNLNKFNITRI 472  
DB 422 AIKRVMASEETEGFRSAKEYKEMAREAI-BEGGSSYNGMATLIQDI 467

## RESULT 10

W47172  
ID W47172 standard; Protein; 470 AA.  
AC W47172;  
DT 08-JUN-1998 (first entry)  
DE Glucosyl transferase (Grase) protein encoded by TW11 gene.  
KW Glucosyl transferase; Grase; TW11; tomato; signalling pathway;  
KW salicylic acid; jasmonic acid; ethylene; wound inducible gene;  
OS Lycopersicon sp.  
PN W09745546-A1.  
PD 04-DEC-1997.  
PF 30-MAY-1997; G01473.  
PR 31-MAY-1996; GB-011420.  
PA (UYO-) UNIV YORK.  
PI Bowles DJ, Calvert CM, ODonnell PJ, Roberts MR;  
DR WPI; 98-032653/03.  
DR N-PSDB; V17054.  
PT Tomato wound inducible (TW11) gene encoding glucosyl transferase -  
PT useful to develop products that alter signalling pathways in plants  
PT by altering of salicylic acid, jasmonic acid or ethylene  
PS Claim 2; Fig 3; 52pp; English.  
CC This is a glucosyl transferase (Grase) protein encoded by a wound  
CC inducible gene (TW11) isolated from wounded tomatoes. The TW11 gene  
CC encodes this Grase from amino acid position 5. The TW11 gene can be  
CC used to identify homologue Grase encoding genes isolated from tobacco  
CC and rice. A microbial host can be transfected or transformed with a  
CC vector containing the Grase encoding nucleic acids. The products can be  
CC used to interfere with Grase and therefore alter signalling pathways in  
CC plants, specifically tobacco, rice or tomato plants by altering levels  
CC of salicylic acid, jasmonic acid or ethylene. This can induce the  
CC production of plant defence proteins such as pathogenesis-related (PR)  
CC and proteinase inhibitor (PIN) proteins which regulate plant development  
CC (plant growth, reproduction and senescence) and improve plant response to  
CC pathogens.  
SQ Sequence 470 AA;

Query Match 15.8%; Score 392.5; DB 1; Length 470;  
Best Local Similarity 26.6%; Pred. No. 3e-30;  
Matches 137; Conservative 75; Mismatches 191; Indels 113; Gaps 21;

QY 12 FP--AQGHINPLEFA-----KRLNTGVVDQVTFETSVYALRRMFETD----- 54  
DB 9 FPDQAQGHMIFTLDMANVACRGVKATITPLNESVFSKAIERKHHGIEIDI 68  
QY 55 -----PSSRIDFVAXXSDYDDGLKGGDKNMYSEMRRKGTALKADTLKLNDAAWG 106  
DB 69 AKENDLPEDCERLDLPVSDDKLPNFK-----AAAMMKDEPEELI----- 108  
QY 107 SECYNRVSVVYSHLFSMAAEVAREVDVPSALLWIEPATVDFVYFYP-----NGYA 158  
DB 109 GEC---RPDCLVSDMFLPWTTSAAKFSIPRIVF-----HGTSYFALCVGTIRNKP 159  
QY 159 DDIDAGSDIOIOLNPLP---QLSKODLPSPFLPSPAPFRTLMKKEFDTLDKPKAK---V 212  
DB 160 KNVSDTETVFPDPFHEIRLTRTQLSPFEQSDTEGTMAPHIKAV-----RESDAKSGV 214  
QY 213 LINTDALETOLKAIDRYELI-----SIGLIPSSIFSDGNDPSSNKKSYGGDLFRKA 266  
DB 215 IFNFSYELESY---VEHYTKVVGKKNWAIGPL--SLCNRDIEDKAERKSSID----- 264  
QY 267 DETYMDWLNKSPSSVVYVSGSLRLPKPOMEIATGLSDTKSPVLWVIRNEEGDSQE 326  
DB 265 EHACLKWLDSKSSIVYVCGFSTADFTTAQOMELAMGLEASGQDFI 324  
QY 327 QAEEKLLSFDRHGTERTLGKIVT--WCSOLDVLTHKSVGCFVTHCGWNSATESLACGVP 385  
DB 325 GFEER-----TKEGLIIRGWAPQSVILDHBAIGAFVTHCGWNSLEGISAGVP 373  
QY 386 VVCFPQWPDGQTNKMIEDYWRSGVRV-----RVNEEGGVVDRREIKRVCSEVIKRSRE- 438



```
Db 374 MVTWPFAEQFFNEKLVTEYMRSGAGVSGQWKRTASEG--VKREATAIKRVMASEET 431
QY 439 --LRESAMMKGLAKEAMDEERGSSNNLNKFNITRI 472
Db 432 EGFERSAKEYKEMAREAI-EGGSSYNGWATLIQDI 466

RESULT 11
ID W64569
AC W64569 standard; Protein; 488 AA.
DE 07-DEC-1998 (first entry)
DT Potato solanidine UDP-glucose glucosyltransferase.
KW Solanidine UDP-glucose glucosyltransferase; SGT; potato; tomato;
  glycoalkaloid; solasodine; solanaceous plant; transgenic plant.
OS Solanum tuberosum cv. Lemhi Russet.
FH Key Location/Qualifiers
FT Region 12..46
FT /note= "conserved domain"
FT Region 110..143
FT /note= "putative steroid binding region"
FT Region 351..401
FT /note= "UDP-glucose binding region"
PN W09834471-AL.
PD 13-AUG-1998.
PF 30-JAN-1998; U01864.
PR 07-FEB-1997; US-797226.
PA (USDA ) US SEC OF AGRIC.
PI Allen PV, Belknap W, Friedman M, Moehs CP, Rockhold DR,
  Stapleton A;
DR WPI; 98-446828/38.
DR N-PSDB; V49609.
PT New DNA encoding solanidine UDP-glucose glucosyl-transferase and
  related vectors - and transgenic plants, used to reduce contents of
  steroidal glyco-alkaloids(s) in plants, specifically potato
  Example 1; Fig 2; 54pp; English.
CC This is the amino acid sequence of potato cv. Lemhi Russett
  solanidine UDP-glucose glucosyltransferase (SGT). It was deduced
  from an SGT cDNA sequence (see V49609). SGT is involved in the
  biosynthesis of steroidal glycoalkaloids in solanaceous plants,
  catalysing the UDP-glucose dependent conversion of the aglycone
  solanidine to gamma-chaconine. The invention relates to DNA
  sequences which encode SGT, and their use, particularly use of an
  antisense DNA construct to inhibit SGT activity, and glycoalkaloid
  levels in solanaceous plants. Transgenic plants are claimed,
  particularly tomato and potato, that have reduced contents of toxic
  steroidal glycoalkaloids. The inactivation of glycoalkaloid
  biosynthetic pathways is beneficial to reduce or eliminate
  glycoalkaloid biosynthesis during storage and shipping.
  Sequence 488 AA;

Query Match 15.4%; Score 382; DB 1; Length 488;
Best Local Similarity 23.5%; Pred. No. 3.4e-29;
Matches 127; Conservative 92; Mismatches 186; Indels 136; Gaps 18;

QY 6 HILLATPPAGHINPSLEFAKRLNTGYVDQVTFFTSVYALRRMRFETDPSRRID----- 60
Db 12 HVLPLPFLSAGHIFPLNARLFASRGV--KATILTPHNALLFRSIDDDVRISGFPIS 69
QY 61 -FVAXXSDYDGLKKGDDGKNYMS--EMRKR-----GTKALKDRTLKLNDAAMGSECY 110
Db 70 IVTIKFPSEAUGLPEGIESFNATSPEPHKIFYALSLLQKPMEDKIRELRPDCIFSDMY 129
QY 111 NRVSFVYVYSHLFSWAAEVAREVDVPSALLWTEPATVDYVYFNGYADDIAGSDQIOL 170
Db 130 -----FPMVTVDIADHLPRIYLNLSAYMCYSIMH-----NLKVR 165
QY 171 PNLPLQSKDQLPFLPSSPARFTLMKEFDTLDKPEPKA----- 211
Db 166 PH-KQPNLDESQSPVPLGPDIEIKFLKSQLTDLRLKSDQDKTVFDELLEQVEDSEERSY 224
QY 212 VLINTFALTEQLKAIDRYELIS-----IGPLI-----PSSIFSDGNDPSSSNKSY 258
```

```
Db 225 IVHDTFVELEPAY---VDYQKLKPKCMHFGPLSHFASKIRSKELISEHNN----- 273
QY 259 GGDLFRAKADETYMDLNSKPSSVVYVYSGSLRLRLPOMEEIAGLSDTKSPVLWIRR 318
Db 274 -----NEIVIDWLNAQPKSVLYVYVSGSMARFPESQNLNEIAQALDASNVFFIVLRP 325
QY 319 NEE-----GDEQEAEEEEKLLSFDFRHGTERLGRIVTWCSQDLVLTHTKSVGCFVTHC 371
Db 326 NEETASWLPVGNLEDK-----TKKGLIKGWVPOLTIMEHSATGGFMTHC 370
QY 372 GWSAIESLACGVVYVCFQWFOGDTNAKME-----DWRSGRVVRVNEEGVY 421
Db 371 GTNSVLEAITFGVPMITWPLYADQFYNEKVVEVRGLGKIDVWNEGIEI-----TGPVI 426
QY 422 DREIKRCVSEVIKSR-----ELRESAMMKGLAKAMDEERGSSNNLNKFNITRINE 475
Db 427 ESAKIREAIERLMSNGSEIINIRDRVMAWSMAQNAATN-EGSSWNNTALIOHKNY 485
QY 476 N 476
Db 486 N 486

RESULT 12
ID W56451
AC W56451 standard; Protein; 473 AA.
DE 06-AUG-1998 (first entry)
DT UDP-glucose:flavonoid 3,5-O-glucosyl transferase amino acid sequence.
KW UDP-glucose:flavonoid 3,5-O-glucosyl transferase; gentian; rose plant;
  blue flower.
OS Unidentified.
PN J10113184-A.
PD 06-MAY-1998.
PF 15-OCT-1996; 272253.
PR 15-OCT-1996; JP-272253.
PA (TOLG ) TOKYO GAS CO LTD.
DR WPI; 98-315473/28.
DR N-PSDB; V23108.
PT Gentian flavonoid glucosyl transferase gene - used to produce
  recombinant rose plants with blue petal(s)
  Claim 1; Pages 6-7; 10pp; Japanese.
CC The present sequence represents a UDP-glucose:flavonoid 3,5-O-glucosyl
  transferase enzyme. The nucleic acid sequence was amplified from
  CC RNA which had been isolated from gentian (sic) petals. The gene is
  CC used to transform rose plants to produce blue flowers.
  Sequence 473 AA;

Query Match 14.2%; Score 351.5; DB 1; Length 473;
Best Local Similarity 25.3%; Pred. No. 3.3e-26;
Matches 133; Conservative 89; Mismatches 194; Indels 109; Gaps 21;

QY 2 VNKRHILLATPPAGHINPSLEFAKRL-----LNTGVQDVQVTFFTSVYALRRMRFETDP 55
Db 1 MKKAEIIVIFPGSHLGSVELAKLAERNEHLSISVI--IIKPNDRKVSNNLKSLSST 58
QY 56 SSRIDFVAXXSDYDGLKKG---DDGKNYMSMKRGTKALKDRTLKLNDAAMGSECYN 111
Db 59 ASRIKVI-----LKQETISLETGTPLFIQKFIESHKAQVRDFL-----AGVSACES 104
QY 112 -RVSFVYVYSHLFSWAAEVAREVDVPSALLWTEPATVDYVYFNGYAD-----DIDAG 164
Db 105 VELSGVYIDMFCTSMIDVANEFEVPSYVFTTSSAAMLGL-WFHFQSLRDNFGKYVDKDS 163
QY 165 SDQQLPNLPLQSKD-LPSPFLPSSPARFTLMKEFDTLD-----KEPKA 211
Db 164 ETVLISIPAFQNLVPGYVLPFI-----FNTEDGCDASLDIOKRFRETKG- 207
QY 212 VLINTFALTEQLKAIDRYELI---SIGPLIPSSIFSDGNDPSSSNKSYGGDLFRKAD 267
Db 208 IINTFLESHAESLSTDETIPVTVGPILGPK-----GSSIESLET 252
```

Qy	285	VSEGSLLRLPKPOMEIAI	TGLSTKSPVLVVR	IRNREGDEQOAESEKLLSFFDRHGTE	344
Db	276	CSFSETFITDQVKALALGLEQTGLPFFLVL	--NFPANVDVSAELNRALPEGF	---LE	329
Qy	345	RL---GKIYV-WCSQLDVLTHK	SVGCFVTHCGHNSAIESI	LACGVPVVCPOWFDOGTNAK	400
Db	330	RVKDKGIHSGWQQQOHI	LAHSSVGCYVCHAGSFSSIEAL	VNDQCVVMIQKGDQTLNAK	389
Qy	401	MIEDVNRSGVRVRNREGGVVDV	RRRETKRCVSEIKSRE	-----LRESAMNKG--LAKE	452
Db	390	LVSGDMEAGVEINRRDEGDFY	GKEDIKEAVEKYVVDVDEKEPG	KLIRENQKKWKEFLNKD	449
Qy	453	AMDERGSSMNL	465		
		:	:	:	
Db	450	IOSKYIGNLVNEM	462		
		:	:	:	

RESULT	14
R26154	
ID	R26154 standard; Protein; 534 AA.
AC	R26154;
DT	27-JAN-1993 (first entry)
DE	HUG-BR2.
KW	Billirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW	monoglucuronide; diglucuronide.
OS	Homo sapiens.
Key	Location/Qualifiers
FT	region
FT	12..22
FT	/note="putative membrane-insertion signal"
FT	region
FT	492..508
FT	/note="putative membrane-anchoring peptide"
FT	modified_site
FT	348
FT	/note="predicted Asn-linked glycosylation site"
FT	misc_difference
FT	282..285
FT	/note="residues encoded by TGCAACGGGAAG 1"
FT	

CC been isolated. They are referred to as HUGRP1 (Q27369) and HUGRP2  
CC (Q27370) (Ritter, et al., J. Biol. Chem. 266:10433-1047 (1991)) and,  
CC upon expression individually in COS-1 cells, encode isoforms that  
CC catalyse the formation of the two bilirubin monoglucuronides and  
CC the diglucuronide.  
CC The cDNAs contain identical 3' ends (1469 bp in length) to each  
CC other and to that of the human phenol transferase cDNA, HLUgp1  
CC (Harding et al., Proc. Natl. Acad. Sci. USA 85:8281 (1988)).  
CC In contrast, they have unique 5' ends.  
SQ Sequence 534 AA;

```

CC      the orthocoronin.
CC      The cDNAs contain identical 3' ends (1469 bp in length) to each
CC      other and to that of the human phenol transferase cDNA, HLU0G1
CC      (Harding et al., Proc. Natl. Acad. Sci. USA 85:8281 (1988)).
CC      In contrast, they have unique 5' ends.
CC      Sequence 534 AA;
SQ
      Query Match      6.6%; Score 164.5; DB 1; Length 534;
      Best Local Similarity 20.5%; Pred. No. 1.le-07;
      Matches 99; Conservative 80; Mismatches 175; Indels 129; Gaps
QY      14 AOGH----INPSLEFAKRLNTGYDVQVTFET--SYVALRRMRPETHDPSSRIDFVAXXDSY 68
      :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      53 ARGHQAVLTPRGEXAHQ-----EERFFTLTAYAV-----PMTQKEPDRVTLGY 96
QY      69 DDG-----LKGDDGKNYMWSEMRRTGKALDKTLIKL--NDAAMGSECYNRVSF--VV 117
      :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      97 TQGFETETHLLKKRYSRMAITNNV----SLAHRCVCVELLHNALL--RHILNATSEFDVL 150
QY      118 YSHLFSWAAREVAREVDVPSALLNIEPATVFDVYVYFYFNGVADDIDAGSDQIQLPN--LPQ 175

```

```
Db 151 TDPNLCGAVLAKYLSIPAVFFW-----RYIPCDLDFKGTQCPNPSSVIPK 196
QY 176 LSK-----QDLPSFLP-----SSPARFTLMKEFTDLDKEPKAKVLINTFD 218
Db 197 LLTNSDHMTFLQVRNMLYPLALSXYCHTFSAPYASLSELF---QREVTVDLLSSAS 253
QY 219 A--LETEOLKAIDRYELISGLIPSSIFSDGNDPSSNKSYGDLFRKADETYMDWLNS 276
Db 254 VWLFRSDFVDYPR-----PIPNMVFVGGINCANG-----KPLSQEFEAYINA 297
QY 277 KPSSVYVVSFGSL--RLPKPQMEETAIGLSDTKSPVLWVIRRNEEGDEQAEKEKLL 335
Db 298 SGEHGIWVSLESWSVSIPEKKAMAIADALGKIPQTVLW----- 336
QY 336 SFDRHCTERLGR-----IVTCSQLDVLTHKSVGCFVTHCGWNSAIESLACGVPVWCF 389
Db 337 ----RYTGRPSNLANNNTILVKLPONDLLGHPMTAFITHAGSHGVYESICNGVPMVM 392
QY 390 POWFDQGTNAKMIEDVNRSGVRVRVNEEGGVDREIKRCVSEVKSRELRESAMMKGL 449
Db 393 PLFGDQMDNAKME--TKGAGVTLNVLE----MTSEDLNAQKAVINDKSYKENIMRLSSL 447
QY 450 AKE 452
Db 448 HKD 450
```

```
CC other and to that of the human phenol transferase cDNA, HLUGPI
CC (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC in contrast, they have unique 5' ends.
SQ Sequence 533 AA;

Query Match 6.3%; Score 156; DB 1; Length 533;
Best Local Similarity 23.4%; Pred. No. 7.3e-07;
Matches 52; Conservative 35; Mismatches 89; Indels 46; Gaps 6;

QY 238 PLIPSSIFSDGNDPSSNKSYGDLFRKADETYMDWLNSKPSSVYVVSFGSL--RLPKP 296
Db 267 PIPNPMVFVGGINCNLHQN-----PLSQEPEAYINASGEHGIWVSLESWSVSIPEK 317
QY 297 QMEETAIGLSDTKSPVLWVIRRNEEGDEQAEKEKLLSFFDRHCTERLGR-----IV 350
Db 318 KAMAIADALGKIPQTVLW-----RYTGRPSNLANNNTILV 352
QY 351 TWCSQLDVLTHKSVGCFVTHCGWNSAIESLACGVPVWCFVQDFQGTNAKMIEDVNRSGV 410
Db 353 KWLFPONDLLGHPMTAFITHAGSHGVYESICNGVPMVMPLFGDQMDNAKME--TKGAGV 411
QY 411 RVRVNEEGGVDREIKRCVSEVKSRELRESAMMKGLAKE 452
Db 412 TLNVLE----MTSEDLNAQKAVINDKSYKENIMRLSSLHKD 449
```

Search completed: August 1, 2000, 19:50:24  
Job time: 21504 sec

```
RESULT 15
R26153
ID R26153 standard; Protein; 533 AA.
AC R26153;
DE HUG-Brl.
KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
OS monoglucuronide; diglucuronide.
FH Homo sapiens.
FT key
FT region 10..20
FT /note= "putative membrane-insertion signal"
FT region 491..507
FT /note= "putative membrane-anchoring peptide"
FT modified_site 102
FT /note= "predicted Asn-linked glycosylation site"
FT modified_site 295
FT /note= "predicted Asn-linked glycosylation site"
FT modified_site 347
FT /note= "predicted Asn-linked glycosylation site"
FT misc_difference 158
FT /note= "feature not labelled in specification"
FT misc_difference 181
FT /note= "feature not labelled in specification"
FT misc_difference 228
FT /note= "feature not labelled in specification"
FT WO9212987-A.
PD 06-AUG-1992.
PE 10-JAN-1992; U00282.
PR 10-JAN-1991; US-639453.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Owens IS, Ritter JK;
DR WPI: 92-284593/34.
DR N-PSDB; Q27369.
PT Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT types I and II
PS Disclosure; Fig 9A-I; 99pp; English.
CC Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC been isolated. They are referred to as HUGBr1 (Q27369) and HUGBr2
CC (Q27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC upon expression individually in COS-1 cells, encode isoforms that
CC catalyse the formation of the two bilirubin monoglucuronides and
CC the diglucuronide.
CC The cDNAs contain identical 3' ends (1469 bp in length) to each
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:51:33 ; Search time 59.57 Seconds  
(without alignments)  
123.004 Million cell updates/sec

Title: US-09-147-955-8  
Perfect score: 2484  
Sequence: 1 MVNKRHILLATFPAQGHNP.....GSSMNLKNFTRIINENAS 478

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues  
Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	631	25.4	471	2	US-08-466-583-2
2	631	25.4	471	4	PCT-US95-07820-2
3	398.5	16.0	471	3	US-09-106-464-2
4	382	15.4	488	2	US-08-797-226-2
5	325.5	13.1	131	2	US-08-466-583-5
6	325.5	13.1	131	4	PCT-US95-07820-5
7	255.5	10.3	131	2	US-08-466-583-6
8	255.5	10.3	131	4	PCT-US95-07820-6
9	205	8.3	63	2	US-08-466-583-8
10	205	8.3	63	4	PCT-US95-07820-8
11	175.5	7.1	534	4	PCT-US92-00282-4
12	170.5	6.9	531	4	PCT-US92-00282-6
13	170	6.8	529	4	PCT-US92-00282-7
14	168	6.8	533	4	PCT-US92-00282-3
15	166	6.7	531	4	PCT-US92-00282-5
16	148.5	6.0	506	5	5180581-2
17	113	4.5	56	2	US-08-466-583-4
18	113	4.5	56	4	PCT-US95-07820-4
19	96.5	3.9	617	1	US-08-361-920-29
20	96.5	3.9	617	1	US-08-479-939-29
21	96.5	3.9	617	1	US-08-483-432-29
22	96.5	3.9	700	2	US-07-862-588B-2
23	96	3.9	906	1	US-08-687-379-4
24	96	3.9	907	1	US-07-718-575-2
25	96	3.9	907	1	US-08-481-206-2
26	96	3.9	907	2	US-08-486-269A-2
27	94.5	3.8	414	2	US-08-750-524-1
28	94.5	3.8	1338	3	US-08-750-141A-3

29	94	3.8	58	2	US-08-466-583-9	Sequence 9, Appli
30	94	3.8	58	4	PCT-US95-07820-9	Sequence 9, Appli
31	92.5	3.7	3052	2	US-08-557-122A-26	Sequence 26, Appl
32	92	3.7	52	2	US-08-466-583-7	Sequence 7, Appli
33	92	3.7	52	4	PCT-US95-07820-7	Sequence 7, Appli
34	91.5	3.7	74	4	PCT-US92-00282-24	Sequence 24, Appl
35	90.5	3.6	1162	2	US-08-599-455B-43	Sequence 12, Appl
36	89.5	3.6	574	2	US-08-907-166-12	Sequence 6, Appli
37	89	3.6	2710	1	US-08-480-604A-6	Sequence 6, Appli
38	89	3.6	2710	2	US-08-405-496A-6	Sequence 5, Appli
39	86.5	3.5	1311	1	US-08-340-011-5	Sequence 2, Appli
40	86	3.5	906	1	US-08-254-573-2	Sequence 2, Appli
41	86	3.5	906	1	US-08-687-379-2	Sequence 2, Appli
42	85.5	3.4	781	1	US-08-373-134D-2	Sequence 2, Appli
43	85.5	3.4	781	2	US-09-114-637-2	Sequence 2, Appli
44	85.5	3.4	828	1	US-08-261-304-2	Sequence 2, Appli
45	85.5	3.4	1479	2	US-08-951-912-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-08-466-583-2  
; Sequence 2, Application US/08466583  
; Patent No. 5919998  
; GENERAL INFORMATION:  
; APPLICANT: Bandurski, Robert S.  
; APPLICANT: Szczylowski, Krzysztof  
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
; TITLE OF INVENTION: and Plant Growth.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee & Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,583  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/265,427  
FILING DATE: 24-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 11-94A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
TELEX: 49617824  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 471 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-466-583-2

Query Match 25.4%; Score 631; DB 2; Length 471;  
Best Local Similarity 33.7%; Pred. No. 2e-56;  
Matches 163; Conservative 78; Mismatches 197; Indels 46; Gaps 16;

```

Qy 6 HILLATPAQGHINPSLEFAKRLNTGYVDQVTFSTSYALRRMRFETDPSRIDFVAXX 65
Db 4 HVLVPPGQGHMNPVQFAKRLASKGVA--TTLVTRFIQRTADVDADHPAM---VEAIS 58
Qy 66 DSYDDGLKGGDDGKNYASEMKRGTALKDPLIKLNDAAAGSECVNRYSFVYVSHLFSWA 125
Db 59 DGHDEG---GFASAAGVAEYLEKQAAAASLASLVEARASSA--DAFTCVVYDSYEDWV 113
Qy 126 AEVAREVDVPSALLWIEPATVDFVYFYFNG-----YADDIDAGSDQIQLP---NL 173
Db 114 LPVARRMGLPAVPFSTQSCAVSAVYHYFHSQGRLLAVPPGAAADSGDGGAGAAALSEAFGL 173
Qy 174 POLSKODLPSPSSPARFETLMEKEDTLDKEPKAK-VLINTFDEALETEQLKAIDRY- 231
Db 174 PEMERSELPSPVDFHGP--YPTIAMQAIKQFAHAGKDDWVLFNSFEETEVLGLTKYL 231
Qy 232 ELISIGPLIPSSIFSDGNDPSSNNK-SYGGDLFRKADTYMDWLNKSPSSVYVSGSL 290
Db 232 KARAIQPCVP--LPTAGRTAGANGRITYGANLV-KPEDACTKWLDTKPDERSVAVYVSGSL 288
Qy 291 LRLPKPOMEETAIGLSDTKSPVLWIRNREGDEQEAEEKEKLLSPFDRHGTER-LGKI 349
Db 289 ASLNAQKEELARGLLAAGKPLWVVR-----ASDEHQVPRYLLABATATGAAMV 338
Qy 350 VTWCSDLDVLTGKSVGCFVTHCGWNSAIESLACGVPVYVCFQWFDQGTNAKMIEDVWRSG 409
Db 339 VPWCPQLDVLHAHPAVGCFVTHCGWNSLTLEALSFGVPVYAMALWTDQPTNARNVELANGAG 398
Qy 410 VVRVNEEGVVDVRRRETKRCYSEVSKRE----LRESAMMKGLAKEAMDEERGSSMNNL 465
Db 399 VRARRDAGAGVFLRGEVERCVRAYMDGGEAASAAKAAAGWRDRARAIV-APGSSSDRNL 457
Qy 466 KNFI 469
Db 458 DEFV 461

```

## RESULT 2

```

PCT-US95/07820-2
; Sequence 2, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089

```

```

; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-07820-2

Query Match 25.4%; Score 631; DB 4; Length 471;
Best Local Similarity 33.7%; Pred. No. 2e-56;
Matches 163; Conservative 78; Mismatches 197; Indels 46; Gaps 16;

Qy 6 HILLATPAQGHINPSLEFAKRLNTGYVDQVTFSTSYALRRMRFETDPSRIDFVAXX 65
Db 4 HVLVPPGQGHMNPVQFAKRLASKGVA--TTLVTRFIQRTADVDADHPAM---VEAIS 58
Qy 66 DSYDDGLKGGDDGKNYASEMKRGTALKDPLIKLNDAAAGSECVNRYSFVYVSHLFSWA 125
Db 59 DGHDEG---GFASAAGVAEYLEKQAAAASLASLVEARASSA--DAFTCVVYDSYEDWV 113
Qy 126 AEVAREVDVPSALLWIEPATVDFVYFYFNG-----YADDIDAGSDQIQLP---NL 173
Db 114 LPVARRMGLPAVPFSTQSCAVSAVYHYFHSQGRLLAVPPGAAADSGDGGAGAAALSEAFGL 173
Qy 174 POLSKODLPSPSSPARFETLMEKEDTLDKEPKAK-VLINTFDEALETEQLKAIDRY- 231
Db 174 PEMERSELPSPVDFHGP--YPTIAMQAIKQFAHAGKDDWVLFNSFEETEVLGLTKYL 231
Qy 232 ELISIGPLIPSSIFSDGNDPSSNNK-SYGGDLFRKADTYMDWLNKSPSSVYVSGSL 290
Db 232 KARAIQPCVP--LPTAGRTAGANGRITYGANLV-KPEDACTKWLDTKPDERSVAVYVSGSL 288
Qy 291 LRLPKPOMEETAIGLSDTKSPVLWIRNREGDEQEAEEKEKLLSPFDRHGTER-LGKI 349
Db 289 ASLNAQKEELARGLLAAGKPLWVVR-----ASDEHQVPRYLLABATATGAAMV 338
Qy 350 VTWCSDLDVLTGKSVGCFVTHCGWNSAIESLACGVPVYVCFQWFDQGTNAKMIEDVWRSG 409
Db 339 VPWCPQLDVLHAHPAVGCFVTHCGWNSLTLEALSFGVPVYAMALWTDQPTNARNVELANGAG 398
Qy 410 VVRVNEEGVVDVRRRETKRCYSEVSKRE----LRESAMMKGLAKEAMDEERGSSMNNL 465
Db 399 VRARRDAGAGVFLRGEVERCVRAYMDGGEAASAAKAAAGWRDRARAIV-APGSSSDRNL 457
Qy 466 KNFI 469
Db 458 DEFV 461

```

## RESULT 3

```

US-09-106-464-2
; Sequence 2, Application US/09106464
; Patent No. 6011145
; GENERAL INFORMATION:
; APPLICANT: Steffens, John C.
; APPLICANT: Ghangas, Gurdev S.
; APPLICANT: Kuai, Jian-ping
; APPLICANT: Eannetta, Nancy
; TITLE OF INVENTION: Chain Length Specific UDP-Glc-Fatty Acid
; TITLE OF INVENTION: Glucosyltransferases
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones, Tullar & Cooper, P.C.
; STREET: P.O. Box 2266 Eads Station
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



Db	371	GTNSVLEATFGVPMITWPLIADQFTNEKVVVRGLGIGIDVWNEGIEI----	---TGPMI	426
QY	422	DREIRKRCVSEVIKSR-----ELRESAMWKGLAKEAMDEERGSSNNLKNFTRI	INE	475
Db	427	ESAKIREATERLMSNGSEEIINIRDVMAKSMQAQNTN-EGGSWNLTALIQHKNY	485	
QY	476	N	476	
Db	486	N	486	

RESULT 5  
US-08-466-583-5  
: Sequence 5, Application US/08466583  
: Patent No. 5919998  
: GENERAL INFORMATION:  
: APPLICANT: Bandurski, Robert S.  
: APPLICANT: Szerszen, Jędrzej B.  
: APPLICANT: Szczygłowski, Krzysztof  
: TITLE OF INVENTION: Genetic Control of Plant Hormone Levels  
: TITLE OF INVENTION: and Plant Growth.  
: NUMBER OF SEQUENCES: 9  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Greenlee & Winner, P.C.  
: STREET: 5370 Manhattan Circle, Suite 201  
: CITY: Boulder  
: STATE: Colorado  
: COUNTRY: United States of America  
: ZIP: 80303

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

FILING DATE: 06-JUN-1993  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/365,427  
 FILING DATE: 24-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ferder, Donna M.  
 REGISTRATION NUMBER: 33,878  
 REFERENCE/DOCKET NUMBER: 11-94A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 303/499-8080  
 TELEFAX: 303/499-8089  
 TELEX: 49617824  
 INFORMATION FOR SEQ ID NO: 5:

```

/ INFORMATION FOR SEQ ID NO: 5:
/
/ SEQUENCE CHARACTERISTICS:
/   LENGTH: 131 amino acids
/   TYPE: amino acid
/   STRANDEDNESS: single
/   TOPOLOGY: linear
/
/ MOLECULE TYPE: protein
/   HYPOTHEetical: NO
/   FRAGMENT TYPE: internal
/
/ US-08-466-583-5

```

Query Match 13.1%; Score 325.5; DB 2; Length 131;  
Best Local Similarity 46.4%; Pred. No. 6.4e-26;  
Matches 64; Conservative 18; Mismatches 45; Indels 11

QY	273	WLNSKPSSVVVYSPGSLRLRPLKPQMBEITAIGLSDTSKPSVLVWVRNNEEGDEQEQAESEE	332
Db	4	WLDTKDRSVVYVYSGSLASGNAQKEELARGLLAAGKPLFVYVR-----ASDEH	53
QY	333	KLLSFDRHGTGR-LGKIYTWCSQLDLVTHKSYGCFVTHCGWNASIAESLACGVPVYVCFQ	391
Db	54	QVPYRLLAENATGAAMVYVWPCPOLDLAHVAGCFVTHCGWNSTLEALSFGVPVYVAMAL	113

```

QY      392 WFDQGTNAKMIEDVWRSG 409
      | | | | | : | | : |
Db      114 WTDOPTNARNVELAWGAG 131

```

```

RESULT      6
PCT-US95-07820-5
: Sequence 5, Application PC/TUS9507820
: GENERAL INFORMATION:
: APPLICANT: Bandurski, Robert S.
: APPLICANT: Szerszen, Jędrzej B.
: APPLICANT: Szczylowski, Krzysztof
: TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
: TITLE OF INVENTION: and Plant Growth.
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee & Winner, P.C.
: STREET: 5370 Manhattan Circle, Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: United States of America
: ZIP: 80303

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/265,427  
FILING DATE: 24-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 11-94B P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
TELEX: 49617824

```

/ INFORMATION FOR SEQ ID NO: 5:
/
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 131 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: NO
/ FRAGMENT TYPE: Internal
/
PCT-US95-07820-5

```

Query Match	13.1%	Score 325.5;	DB 4;	Length 131;
Best Local Similarity	46.4%;	Pred. No. 6.4e-26;		
Matches 64;	Conservative	18;	Mismatches 45;	Gaps 2;
			Indels 11;	

[illegible]

## RESULT 7



```

US-08-466-583-6
; Sequence 6, Application US/08466583
; Patent No. 591998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerzen, Jedrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: Internal
US-08-466-583-6

Query Match          10.3%; Score 255.5; DB 2; Length 131
Best Local Similarity 40.0%; Pred. No. 1e-18;
Matches 56; Conservative 19; Mismatches 50; Indels 1

QY   273 WLNSKPSSVVVVSFGSLRLRPK-QMEETAIGLSDTSKSPVLWVIRNNEBGDSQE
      || | ||||| : : - : || : || | | : || | :
Db   4  WLGRQPARGVAYVSFGT-VACPRPDELRELAAGLSDGAPFLSLR-----
QY   332 EKLL--SFEDRHGTERLGKTVTCWSOLDVLTKHSVCFTVHCWNSAIESLAGCVI
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   52 WPHLPGLDRAAGTSGLVVPWPAPQVAVLRHPBSVGAFTVHAGWASVLEGLSSGV
      | : || || : | : |
QY   390 POWFDQGTNAKMIEDYWRSG 409
      | : || || : | : |
Db   112 PFPGDQMNRARSVAHWVGFG 131

RESULT      8
PCT-US95-07820-6
; Sequence 6, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerzen, Jedrzej B.
; APPLICANT: Szczygłowski, Krzysztof

```

```

; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: Internal
; PCT-US95-07820-6

Query Match 10.3%; Score 255.5; DB 4; Length 131
Best Local Similarity 40.0%; Pred. NO. 1e-18;
Matches 56; Conservative 19; Mismatches 50; Indels 1

QY 273 WLNSKPSSVVVYFSGSLLELPKP-QMEETAIAGLSKTSPLVWVIRNBSGDQOE
Db 4 WLGRQARGVAYVYFSGT-VACPRDELRELAEGLSDSGAPFLMSLR-----
QY 332 EKLL--SFDRDHGTERLGKIVTWCSOLDVLTHKSVGCFVTHCGWNSAIESLAGV
Db 52 WPHLPFGFLDRAAGTSGSLVWPWAPQVAVLRHFSVGAFTVHAGWASVLEGLSSGV
QY 390 POWFDQGTNAKMTIEDVWRSG 409
Db 112 PPFGDQRMNARSVAHVWGFG 131

RESULT 9
US-08-466-583-8
; Sequence 8, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szezszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201

```

```

; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: Internal
; PCT-US95-07820-6

Query Match 10.3%; Score 255.5; DB 4; Length 131
Best Local Similarity 40.0%; Pred. NO. 1e-18;
Matches 56; Conservative 19; Mismatches 50; Indels 1

QY 273 WLNSKPSSVVVYFSGSLLELPKP-QMEETAIAGLSKTSPLVWVIRNBSGDQOE
Db 4 WLGRQARGVAYVYFSGT-VACPRDELRELAEGLSDSGAPFLMSLR-----
QY 332 EKLL--SFDRDHGTERLGKIVTWCSOLDVLTHKSVGCFVTHCGWNSAIESLAGV
Db 52 WPHLPFGFLDRAAGTSGSLVWPWAPQVAVLRHFSVGAFTVHAGWASVLEGLSSGV
QY 390 POWFDQGTNAKMTIEDVWRSG 409
Db 112 PPFQDQRMNARSVAHVWGFG 131

RESULT 9
US-08-466-583-8
; Sequence 8, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szezszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201

```







QY 238 PLIPSSIFSDGNDPSSSNKSYGGDLFRKADETYMDWLNKPESSVYVVSFGSLL-RLPKP 296  
Db 265 PVMPNVFIFGG-----INCKRRDL-----SQEPEAYINASGEHGIIVVFSIGSNVSEIPEK 315  
QY 297 QMEETIAIGLSDTKSPVLWVIRNEEGDEQAEKEKLLSFFDRHGCTERLGK-----IV 350  
Db 316 KAWAIADALGKNPQTVLW-----RYTGTSPSNLANNTILV 350  
QY 351 TWCSQLDVLTHKSVGCFVTHCGWNSAIESLACGVPVVCFPQWFDQGTNAKMIEDVWRSGV 410  
Db 351 KWL PQNDLLGHPTNTRAFITHAGSHGYESICNGVPVWMPPLFGDQMDNAKME-TKGAGV 409  
QY 411 RVRVNEEGVVDREIKRCVSEVIKSRRELRESAMMKGLAKE 452  
Db 410 TLNVLE----MTSEDLLENALKAVINDKSYKENIMRLSSLHKD 447

Search completed: August 1, 2000, 19:51:34  
Job time: 18756 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:32:19 ; Search time 31.57 seconds  
(without alignments)  
469.205 Million cell updates/sec

Title: US-09-147-955-8

Perfect score: 2484

Sequence: 1 MVNKRHILLATFPAQGHNP.....GSSMNNLKNFTIRINENAS 478

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	631	25.4	471	1	IRAG_MAIZE
2	453	18.2	453	1	UFOG_GENTR
3	420	16.9	433	1	UFOG_SOLME
4	382	15.4	454	1	ZOX_PRAVU
5	373	15.0	459	1	ZOG_PHALU
6	357.5	14.4	449	1	UFO1_MANES
7	354	14.3	487	1	UFO5_MANES
8	349.5	14.1	346	1	UFO2_MANES
9	340	13.7	287	1	UFO7_MANES
10	326.5	13.1	471	1	UFO3_MAIZE
11	325.5	13.1	471	1	UFO1_MAIZE
12	324	13.0	473	1	UFOG_PETHY
13	323.5	13.0	394	1	UFO6_MANES
14	323.5	13.0	471	1	UFO2_MAIZE
15	313.5	12.6	455	1	UFOG_HORVU
16	281.5	11.3	241	1	UFO4_MANES
17	240	9.7	154	1	UFOG_VITVI
18	184.5	7.4	535	1	UD11_RAT
19	182	7.3	535	1	UD11_MOUSE
20	175.5	7.1	534	1	UD14_HUMAN
21	174.5	7.0	533	1	UD12_MOUSE
22	172.5	6.9	532	1	UD14_RABIT
23	170.5	6.9	533	1	UD12_RAT
24	170.5	6.9	534	1	UD15_HUMAN
25	170	6.8	529	1	UD16_RAT
26	170	6.8	531	1	UD16_MOUSE
27	169	6.8	530	1	UD18_RAT
28	167.5	6.7	534	1	UD13_HUMAN
29	167	6.7	533	1	UD11_HUMAN
30	166.5	6.7	531	1	UD16_RABIT
31	166	6.7	531	1	UD16_HUMAN
32	165.5	6.7	520	1	UD17_MOUSE
33	163	6.6	530	1	UD12_HUMAN

34	163	6.6	531	1	UD13_RAT	Q64637 rattus norv
35	161.5	6.5	531	1	UD15_RAT	Q64638 rattus norv
36	160.5	6.5	531	1	UGTE_CABEL	Q10941 caenorhabdi
37	160	6.4	530	1	UDBH_WACFA	Q77649 macaca fasc
38	160	6.4	541	1	CGT_HUMAN	Q16880 homo sapien
39	155	6.2	531	1	UD17_RAT	Q64633 rattus norv
40	154	6.2	488	1	UDPE_NPVLD	P41713 lymantria d
41	154	6.2	541	1	CGT_MOUSE	Q64676 mus musculu
42	154	6.2	541	1	CGT_RAT	Q09426 rattus norv
43	153.5	6.2	491	1	UDPE_NPVCF	Q90157 choristoneu
44	153.5	6.2	529	1	UDB7_HUMAN	P16662 homo sapien
45	153	6.2	430	1	OLED_STRAT	Q53685 streptomyce

#### ALIGNMENTS

```

RESULT 1
IAAG_MAIZE
ID IAAG_MAIZE STANDARD; PRT; 471 AA.
AC Q41819;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE INDOL-3-ACETATE BETA-GLUCOSYLTRANSFERASE (EC 2.4.1.121) (IAA-GLU
DE SYNTHETASE) ((URIDINE 5'-DIPHOSPHATE-GLUCOSE:INDOL-3-YLACETYL)-BETA-D-
DE GLUCOSYL TRANSFERASE).
GN IAGLU.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.
RC TISSUE-ENDOSPERM;
RX MEDLINE; 94367368.
RA Szeleszen J.B., Szczyglowski K., Bandurski R.S.;
RT *Iaglu, a gene from Zea mays involved in conjugation of growth
RT hormone indole-3-acetic acid."
RL Science 265:1699-1701(1994).
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + INDOL-3-ACETATE = UDP + INDOL-3-
CC 3-ACETYL-BETA-1-D-GLUCOSE.
CC -1- PATHWAY: FIRST STEP IN INDOL-3-ACETIC ACID CONJUGATION PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLUCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L34847; AAA59054.1; -.
DR MA12EDB; 83603; -.
DR PFAM; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; glycosyltransferase.
FT DOMAIN 80 85 POLY-ALA.
SQ SEQUENCE 471 AA; 49710 MW; 99A7F56C2BB4EB39 CRC64;

```

Query Match 25.4%; Score 631; DB 1; Length 471;  
Best Local Similarity 33.7%; Pred. No. 1.6e-39;  
Matches 163; Conservative 78; Mismatches 197; Indels 46; Gaps 16;

Qy	6	HILLATFPAQGHNP	SLFAKRLNLTGYDVTFVSVVLRMRFPETDPSRRIDFVAXX	65
Db	4	HVLVFPFGQGHMNP	VQFAKRLASKGVA--TTLVTRFRIQRTADVDAHPAM--VEAIS	58
Qy	66	DSYDGLKKGDDG	GKNYMSMKRGTGKLTALDKTLKLNDAAMGSECYNRVSVVYSHLFSWA	125
Db	59	DGHDEG---	GFASAAGVAEYLEKQAAASASLASLVEARASSA--DAFTCVYDSYEDWV	113

```

QY 126 AEVAREVDVPSALLWIEPATVDFVYVYFNG-----YADDIDAGSDQIQLP-----NL 173
Db 114 LPVARRMGLPAVPSTOSCAVSYVYHFSQGR LAVPFGAAGSDGCGAGAALSEAFGL 173
QY 174 POLSKQDLPSLPSPARFRTLMKEKFDTLDKPEKAK-VLINTFDALTEQLKAIDRY- 231
Db 174 PEMERSELPSVFDRGP--YPTIAQAIKQPAHAGKDDWVLFNSFEELETVLAGLTKYL 231
QY 232 ELISIGPLIPSSIFSDGNDPSSNK-SYGGDLFRKADETYMDWLNKSPSSVYVYSGSL 290
Db 232 KARAIGFCVP--LPTAGTAGANRITYGANLV-KPEDACTKWLDTPKDRSVAVYSGSL 288
QY 291 LRLPKPQMEETAIGLSDTKSPVLVIRNREGDEQAEKEKLLSFDFRGTGER-LGKI 349
Db 289 ASLNAQKEELARGLLAAGKPFLLVVR-----ASDEHQVPRVLLAATATGAAMV 338
QY 350 VTWCSQDLVTHKSGVCFVTHCGWNSAIESIACGVVPVCFQWFDQGTNAKMIEDVWRSG 409
Db 339 VPMCPQLDVLHAHPAVGCFVTHCGWNSLEALSFGVPMVAMALWTDQPTNARNVELAWGAG 398
QY 410 VVRVNEEGGVDRREIKRCVSEVIKSPRE-----LRESAMMKGLAKEMDEERGSSHNLL 465
Db 399 VRARRDAGVFLRGEVERCVRAYMDGGEAASAARKAAGEWRDRARAAY-APGSSDRNL 457
QY 466 KNFI 469
Db 458 DEFV 461

RESULT 2
UFOG_GENTR STANDARD; PRT; 453 AA.
AC Q96493;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE 3-O-GLUCOSYLTRANSFERASE).
OS Gentiana triflora.
OC Eukaryota; Viridiplantae; Tracheophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Gentianales;
OC Gentianaceae; Gentiana.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PETAL;
RX MEDLINE; 96416435.
RA Tanaka Y., Yonekura K., Fukuchi-Mizutani M., Fukui Y.,
RA Fujiwara H., Ashikari T., Kusumi T.;
RT "Molecular and biochemical characterization of three anthocyanin
RT synthetic enzymes from Gentiana triflora.";
RL Plant Cell Physiol. 37:711-716(1996).
CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D85186; BAAL2737.1; -.
DR PFAM; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW transferase; Glycosyltransferase.
SQ SEQUENCE 453 AA; 50009 MW; BF738B0A2DA76C05 CRC64;

```

```

Query Match 18.28; Score 453; DB 1; Length 453;
Best Local Similarity 26.6%; Pred. No. 2.4e-26;
Matches 137; Conservative 90; Mismatches 162; Indels 126; Gaps 19;

QY 6 HILLATPAQOHINPSLEFAKRLNTGVVDQVTFP-----TSVVALRRM----- 49
Db 6 HVAVLAPFPFGTHAAPLLTLVNRLAASAPDIIFSFSSSSITTFPSPTNLISGNIKPY 65
QY 50 -----RFETDPSRIDFV--AXXDSYDDGLKKGDDGKNYMSMKRGTGRTKALKDPTL 97
Db 66 AVWDGSPGFGVFCGNPREPIEYFLNAAFDNFKAM-----KAVEDTG 108
QY 98 IKLNDAMGSCYNRVSVFYVYSHLFSW-AAEVAREVDVPSALLWIEPATVDFVYVYFNG 156
Db 109 VNIS-----CLLTDAFLWFAADFSEKIGVPWIPVWVTAASCSCSLHV----- 149
QY 157 YADDIDAGSDQIQ-----LPNLPOLSKQDLR-SFLPLSPSPARFRTLMKEKFDTL 204
Db 150 YTDEIRSFRAEFDIAEKAETIDIFPGLSAISFSDLPBELIMEDSQSIFALTLMNGLKL 209
QY 205 DKEPKAKVLINTFDAL-----TEQLKAIDRYELISIGPL--IPSSIFSDGNDPSSNKS 259
Db 210 HK--ATAVAVNSFEEDIPITNHLRSTNQLNLTGIPLOTLSSTIPPEDNE----- 258
QY 260 GDLFRKADETYMDWLNKSPSSVYVYSGSLRLPKPQMEETAIGLSDTKSPVLVIRN 319
Db 259 -----CLKWLQTKQESSVYVYSGTINPPNEMAALASTLESKRKIPFLWSLR-- 306
QY 320 EEGDEQEAEEKLL--SFFDRHGTGRIGKIVTWCSQDLVTHKSGVCFVTHCGWNSAI 377
Db 307 -----DEARKHLPENFIDR--TSTFGKIVSWAPQLHVLENPAIGVFFVTHCGWNS 355
QY 378 ESLACGVVPVCFQWFDQGTNAKMIEDVWRSGVRVYVNEEGGVDRREIKRCVSEVI--- 434
Db 356 ESICRVPIGCRPFQDGQKVNARVVEDVWKIGVGK-----GGVTFDETRVLELVLFSD 411
QY 435 KSRELRESNMMWKGLAKEMDEERGSSMNNLNKFI 469
Db 412 KGKEMRQNVGLRKEKAKDAV-KANGSSTRNFESLL 445

RESULT 3
UFOG_SOLME STANDARD; PRT; 433 AA.
AC Q43641;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE 3-O-D-GLUCOSYLTRANSFERASE).
OS Solanum melongena (Eggplant) (Aubergine).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
OC Solanaceae; Solanum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. SINSADOHARANASU; TISSUE-HYPOCOTYL;
RA Toquri T.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```







QY 169 QLPNLPOLSKQDLPSPFLPSSPARFTLMKEKFTDITLKEPK-----AK-VLINTFDALETE 223  
Db 159 QVPGI-----VNSPPSKAMPTA-----ILSKQWFPPLLENTRRYGAAGVLIINTFFLESH 209  
QY 224 QLKADRYELISIGLIPISFSDGNDPSSNKSYGGLDFRKADETYMDLNSKPKSESSV 283  
Db 210 AIESKFPDPIYPVGPIL-----DVRNSG-----RNTNQEIQWLDQDPSSV 252  
QY 284 VVSFGLLRLPKPOMEIEAIGLSDTKSPVLWIRNNEEGDEQOEAEKEKLL-----SFF 338  
Db 253 FLCFSGNSGFSQDKQKEIACALSDSHRFLMSLADRAPGFLFLESQDYEDLQAEVLPEGL 312  
QY 339 DRHGTERLCKIVTWCSQDVLTHKSGVCFTHCGWNSAIESLACGVPVVCFFQWFDQGN 398  
Db 313 ER--TSGIEKVGAPQAVLAHPATGGLVSHSGWNSILESIFWGPVATWPMYAEQQN 370  
QY 399 AKMIEDVRSRGVVRV-----NEEGVDVDRREIKRCVSEVIKSRRELRESAMMKGLAKRA 453  
Db 371 A--FQWVIELGLAVEIKDYRNDSGEIVKCDQIERGIRCLMKHDSRRKKV-----KEM 422  
QY 454 MDEERGSSMNN-----LANFTRII 473  
Db 423 SEKRGALMEGGSSYCWLDNLKDMI 448

RESULT 7  
UFO5\_MANES  
ID UFO5\_MANES STANDARD; PRT; 487 AA.  
AC Q40287;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 5 (EC 2.4.1.91) (UDP-GLUCOSE  
FLAVONOID 3-O-GLUCOSYLTRANSFERASE 5).  
GN GT5 OR UGT73A5.  
OS Manihot esculenta (Cassava) (Manioc).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;  
OC Euphorbiaceae; Manihot.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COTYLEDON;  
RX MEDLINE; 95201291.  
RA Hughes J., Hughes M.A.;  
RT \*Multiple secondary plant product UDP-glucose glucosyltransferase  
genes expressed in cassava (Manihot esculenta Crantz) cotyledons.\*;  
RL DNA Seq. 5:41-49(1994).  
CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS  
GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN  
PIGMENTS (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL  
3-O-D-GLUCOSIDE.  
CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE  
ANTHOCYANIN BIOSYNTHETIC PATHWAY.  
CC -!- TISSUE SPECIFICITY: FAINTLY EXPRESSED IN COTYLEDONS.  
CC -!- DEVELOPMENTAL STAGE: RARE TRANSCRIPTS EXPRESSED IN COTYLEDON AND  
ROOTS DURING THE DEVELOPMENT.  
CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; X77462; CAA54612.1; -.  
CC PFAM; PF00201; UDPGT; 1.  
CC PROSITE; PS00375; UDPGT; 1.  
KW Transferase; Glycosyltransferase; Multigene family.  
SQ SEQUENCE 487 AA; 54381 MW; 0F5CA2EBC697F124 CRC64;

Query Match 14.3%; Score 354; DB 1; Length 487;  
Best Local Similarity 24.6%; Pred. No. 5.7e-19;  
Matches 134; Conservative 88; Mismatches 171; Indels 152; Gaps 22;  
QY 3 NKRHILLATFPAGHINPSLEFAKRLNLTGYVDQVTF-----TSVYALRRMFEFDP-- 55  
Db 8 SRPHIVLLSPGLGHLIPVLELGRIVTLNCFD-VTFWVGSDTSAAEQVLSAMTPKL 66  
QY 56 -----SSRID-----FVAXXDSYDDGLKKGGDKGNKYMSEMRKRGTKALKD 95  
Db 67 CEIIQLPPNISCLIDPEATVCTRLFV-----LMREIRPAFAAAV-- 106  
QY 96 TLIKLNDAAAGSECYRNVSVFVYSHLFSWAAEVARVDVP-----SALLWIEPATVFDVY 150  
Db 107 SALKFRPAAIIVDLFGTESL-----EVAKELGIKAKYVYIASNAWFLALTYI-- 152  
QY 151 YFYFNGYADDIDAGSDQIQLNLPOLSKQDLPSPFLPSS-----PARFRTLMKEKFD 202  
Db 153 -----VPILDKEVEGEFVLQKEPKIPGCRPYRTEEVDPMLD 190  
QY 203 TLDKE-----PKAK-VLINTFDAL-----TEOLKAIDRYELISIGLIP 241  
Db 191 RTNQOYSEYFRLGIEIPTADGILMTWEALEPTTFGALRDVKFLGRVAKVPVFPPIGLRR 250  
QY 242 SIFSDGNDPSSNKSYGGLDFRKADETYMDLNSKPESVSVVYVSGSLRLPKPOMEI 301  
Db 251 QA-----GPCGSNCE-----LLDWLDQOPKESVSVVYVSGSGTLLSLEOMIEL 292  
QY 302 AIGLSDTKSPVLWIRRN--EEGDEQ--EQAEKEEKLSPFFDRHGTREL---GKIV-TWC 353  
Db 293 ANGLERSQORFIVWVROPTVKTGDAAFFTGDDADMSGVFPEGFLTRIONVGLVVPQWS 352  
QY 354 SOLDVITHKSVGCFVTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKMIEDVRSRGVVR 413  
Db 353 PQTHIMSHPSVGVFLSHCGWNSVLESTAGVPIIAWPIYAEQRNATLLTEELGVAVRPK 412  
QY 414 VNEEGGVDRREIKRCVSEVI---KSRRELRESAMMKGLAKEAMDEERGSSMNLKNFIT 470  
Db 413 NLPAKEVVRKEIERMIRRMVDEGSEIKRVRRELKDSGEKALN-EGGSSF-----NYS 467  
QY 471 RIINE 475  
Db 468 ALGNE 472  
RESULT 8  
UFO2\_MANES  
ID UFO2\_MANES STANDARD; PRT; 346 AA.  
AC Q40285;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 2 (EC 2.4.1.91) (UDP-GLUCOSE  
FLAVONOID 3-O-GLUCOSYLTRANSFERASE 2) (FRAGMENT).  
GN GT2 OR UGT73A2.  
OS Manihot esculenta (Cassava) (Manioc).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;  
OC Euphorbiaceae; Manihot.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COTYLEDON;  
RX MEDLINE; 95201291.  
RA Hughes J., Hughes M.A.;  
RT \*Multiple secondary plant product UDP-glucose glucosyltransferase  
genes expressed in cassava (Manihot esculenta Crantz) cotyledons.\*;  
RL DNA Seq. 5:41-49(1994).  
CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS  
GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN  
PIGMENTS (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL

CC 3-O-D-GLUCOSIDE.  
 CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE  
 CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN COTYLEDONS, ROOTS AND LEAVES.  
 CC -1- DEVELOPMENTAL STAGE: MAXIMUM EXPRESSION IN COTYLEDONS THAT JUST  
 CC EMERGED FROM THE SEED COAT. LOW LEVELS IN HYPOCOTYLS AND  
 CC INCREASING LEVELS IN ROOTS THROUGHOUT THIS PERIOD OF DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X77461; CAAS4611.1; -  
 CC PFAM; PF00201; UDPGT; 1.  
 CC PROSITE; PS00375; UDPGT; 1.  
 CC TRANSFERASE; Glycosyltransferase; Multigene family.  
 CC NON\_TER 1  
 CC SEQUENCE 346 AA; 38834 MW; 89D51FDF8E0D5F2 CRC64;

Query Match 14.1%; Score 349.5; DB 1; Length 346;  
 Best Local Similarity 29.3%; Pred. No. 7.7e-19;  
 Matches 111; Conservative 70; Mismatches 117; Indels 81; Gaps 17;

QY 127 EVAREVDVPSALLWIEPATVDFVYFYFNG-----YADDI-----DAGSD 166  
 DB 7 DLADDFGIPS-----YIFFASGGGFLGFMLYVOKIHDEENFNPIEFKSDSTE 53  
 QY 167 QIQLPNLPOLSKDLPFLPSPARPFTLMKEFDIL----DREPKAK-VLINTFDAL 221  
 DB 54 LI-VPSEL-----VNPFFRILPSS-----ILNKERFGQLLAIAKFRQAQGIIVNTFLE 103  
 QY 222 TEQLKATDRVELISIGPLIPSSIFSDGNDPSSNKSYYGDLFRKADETYMDLNSKPRESS 281  
 DB 104 SRAIESKVPPLHYVHGPII--DVKSDG-----RTHPEIMQWLDQPEGS 146  
 QY 282 VVYVYVFGSLRLPKPOMEETAIGLSDTKSPVLWVIRNEEGDE---OEQAEEEEKLL--S 336  
 DB 147 VVFLCFGSMGFSBDQLKATAYALENSGHRFLWSIRPPDPDKIASPTDYEDPRDVLPEG 206  
 QY 337 FFDHGHGTERLGIKIVTWCSDQLDVLTHKSVGCVFTHCGNWSAIESLACGVPVYVCFQWFDQG 396  
 DB 207 FLER--TVAVGKVIWGAQVAVLAHRAHGAIGGVFVSHCGNWSVLESWFGVPIATWPMYABQQ 264  
 QY 397 TNAKMIEDVWRSGVVRVNV-----EEGVVDRREIKRCVSEVIK--SRELRESAMMKVGLA 450  
 DB 265 FNA--FEMVVELGLGVDEIDMGYRKESGIIVNSDKIERAIRKLMENSEDKRKVKEMREKS 322  
 QY 451 KEAMDEERGSGSMNKLKFI 469  
 DB 323 KMAL-IDGGSSFISLGDPI 340

RESULT 9  
 UF07\_MANES STANDARD; PRT; 287 AA.  
 AC Q40289;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 7 (EC 2.4.1.91) (UDP-GLUCOSE  
 DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 7) (FRAGMENT).  
 GN GT7 OR UGT73A7.  
 OS Manihot esculenta (Cassava) (Manioc).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;  
 OC Euphorbiaceae; Manihot.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE-COTYLEDON;  
 RX MEDLINE; 95201291.  
 RA Hughes J., Hughes M.A.;  
 RT \*Multiple secondary plant product UDP-glucose glucosyltransferase  
 RT genes expressed in cassava (Manihot esculenta Crantz) cotyledons.\*;  
 RL DNA Seq. 5:41-49(1994).  
 CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS  
 CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN  
 CC PIGMENTS (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL  
 CC 3-O-D-GLUCOSIDE.  
 CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE  
 CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN COTYLEDONS, HYPOCOTYLS, ROOTS AND  
 CC LEAVES.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN ALL TISSUES (COTYLEDON,  
 CC HYPOCOTYLS AND ROOTS) AT UNIFORM LEVELS AT ALL STAGES OF  
 CC DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X77464; CAAS4614.1; -  
 CC PFAM; PF00201; UDPGT; 1.  
 CC PROSITE; PS00375; UDPGT; 1.  
 CC TRANSFERASE; Glycosyltransferase; Multigene family.  
 CC NON\_TER 1  
 CC SEQUENCE 287 AA; 31397 MW; 0A0D9C153C193571 CRC64;

Query Match 13.7%; Score 340; DB 1; Length 287;  
 Best Local Similarity 28.5%; Pred. No. 3e-18;  
 Matches 92; Conservative 67; Mismatches 96; Indels 68; Gaps 14;

QY 170 LPNLPOLSKDLPFLPSPARPFTLMKEFDTLDE-PR-AKVILINTFDALET----- 222  
 DB 5 IPGMSKIQIRDLPEGVLFGN--LESLSQMLHNMGRMLPRAAAVLMNSFEELDPTIVSD 61  
 QY 223 -----EQLKATDRVELISIGPLIPSSIFSDGNDPSSNKSYYGDLFRKADETYMDLNSK 277  
 DB 62 LNSKFNNILCIGPNLVSPPPVPD-----TYG-----CMAWLDQK 97  
 QY 278 PESSVYVYVFGSLRLPKPOMEETAIGLSDTKSPVLWVIRNEEGDEOEQAEEEEKL--- 334  
 DB 98 KPASVATISFGSVATPPPHLEVALAEAEASKVPFLSLK-----DHSKVHLP 145  
 QY 335 LSFFDRHGHGTERLGIKIVTWCSDQLDVLTHKSVGCVFTHCGNWSAIESLACGVPVYVCFQWFD 394  
 DB 146 NGFLDR--TKSHGIVLSWAPQVEILEHAALGVFVTHCGNWSILESIVGVPMICRPFPGD 203  
 QY 395 QGTNAKMIEDVWRSGVVRVNVNEEGVVDREIKRCVSEVI---KSRELRESAMMKVGLAK 451  
 DB 204 QRLNGRMVEDYWEIGLLM-----DGGVLTKNGAIDGLNOLLOLQGGKKMKRENIRKLKELAK 259  
 QY 452 EAMDEERGSGSMNKLKFI 474  
 DB 260 GA-TEPKGSSS---KSF-TELAN 277

RESULT 10  
 UF03\_MAIZE STANDARD; PRT; 471 AA.  
 ID UF03\_MAIZE  
 AC P16167;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)



Db 13 HAVVAFPPSHAAVLISARALAAAAAPSGATLSFLSTASSLAQLRKASSASAGHGLPG 72  
 QY 58 RIDFVAXXSDYDGLKGGDGC-----KNYSEMRKRGTKALDQTLKLKNDAAAGWSECY 110  
 Db 73 NLRFV-----EVPDCAAAEETVPVPRQMLPWEAAEAGGVKA-----WLEAARAAGGA-- 122  
 QY 111 NRVSFVYVSHLFSW-AAEAREVDVPSALLWIEPATVFDVYFYFNGYADDI-DAGSDQ- 167  
 Db 123 -RVTCVV-GDAFVWPAADAAASAGAPVVPVWTAASCAL-LAHIRTDALREDVDGQAANRV 179  
 QY 168 ----IQPNLPOLSKQDLPSLELSSPARFRLMKKEPFDTLDEKPKAVLINTDALETE 223  
 Db 180 DGLLISHPLGLASYVRVLDPCVWGDFENVYINLLVHRMGQCLPRSAARAVALNTPGLDPP 239  
 QY 224 QLKALDRVELISGLPLPSI-----FSDGNDPSSSKSYGGDLFRKADETYMDWL 274  
 Db 240 DVTA-----ALAEILPNCVFPFGYHLLLAEDDADTAAPADPHG-----CLAWL 282  
 QY 275 NSKPESVYVYFSGSLRLRKP-QMEITAIGLSDTKSPVLWVIRNNEEGDEQAEAEK 333  
 Db 283 GRQPARGVAVYSFCT-VACPRDELRELAAGLEDGAPFLSLR-----EDSWP 330  
 QY 334 LL--SFDRHGTERTGLKIVWCSQDLVTHKSGVCGFVTHCGWNSAIESLACGVVPCPPQ 391  
 Db 331 HLPFCFLDRAAGTSGGLVVPWAPQVAVLRHPSPVGAFVTHAGWASVLEGLSSGVPMACRPF 390  
 QY 392 WFDQGTAKMTEDYWRSGVRVNEEGVDREIKRCVSEVIKRE---LRESAMWKG 448  
 Db 391 FGDORMNARSVAHVWFGAAFE-----GAMTSAGVATAVEELLRGEAGMARAKELQA 445  
 QY 449 LAKEAMDEERGSSMNLKNFI 469  
 Db 446 LVAEAFG-PGGECKRNFDEV 465

## RESULT 12

UFOG\_PETHY STANDARD; PRT; 473 AA.  
 ID UFOG\_PETHY STANDARD; PRT; 473 AA.  
 AC Q43716; Q40895;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID  
 DE 3-O-GLUCOSYLTRANSFERASE) (ANTHOCYANIN RHAMNOSYL TRANSFERASE).  
 GN RT OR UGT72.  
 OS Petunia hybrida (Petunia).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;  
 OC Solanaceae; Petunia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. V-26; TISSUE=LEAF;  
 RX MEDLINE; 9417177.  
 RA Kroon J., Souer E., de Graaff A., Xue Y., Mol J., Koes R.;  
 RT \*Cloning and structural analysis of the anthocyanin pigmentation  
 RT locus Rt of Petunia hybrida: characterization of insertion sequences  
 RT in two mutant alleles\*;  
 RL Plant J. 5:69-80(1994).  
 RN [2]  
 RP SEQUENCE OF 3-473 FROM N.A.  
 RC STRAIN-CV. OLD GLORY BLUE; TISSUE=PETAL;  
 RX MEDLINE; 9417178.  
 RA Brugniera F., Holton T.A., Stevenson T.W., Farcy E., Lu C.Y.,  
 RA Cornish E.C.;  
 RT \*Isolation and characterization of a cDNA clone corresponding to the  
 RT Rt locus of Petunia hybrida\*;  
 RL Plant J. 5:81-92(1994).  
 CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS  
 CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN  
 CC PIGMENTS.  
 CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL  
 CC 3-O-D-GLUCOSIDE.  
 CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE

CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.  
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 DR EMBL; X71060; CAA50377.1; -;  
 DR EMBL; X71059; CAA50376.1; -;  
 DR EMBL; Z25802; CAA81057.1; -;  
 DR PFAM; PF00201; UDPGT; 1;  
 DR PROSITE; PS00375; UDPGT; 1;  
 KW Transferase; Glycosyltransferase.  
 FT CONFLICT 246 255 LIGVVPDPP -> SNRTRSSGPA (IN REF. 1).  
 FT CONFLICT 348 348 N -> H (IN REF. 2).  
 FT CONFLICT 431 431 D -> E (IN REF. 2).  
 SQ SEQUENCE 473 AA; 52405 MW; A310E4BFC9EB02A CRC64;  
 -----  
 Query Match 13.0%; Score 324; DB 1; Length 473;  
 Best Local Similarity 24.3%; Pred. No. 9.1e-17;  
 Matches 120; Conservative 82; Mismatches 217; Indels 74; Gaps 16;  
 Qy 6 HILLATPAAQHINPSLEFAKRLNTGYDQVTFSTYALRRMRFETDPSRIDFVAXX 65  
 Db 13 HVVAFPPFAFHISPFVQANKL--SSYGVKVSFFETASGNASRVKSLMSAPTTHIVPLT 70  
 Qy 66 DSYDDGLKKGDDGKNYMSEMRKGTALKOTLIKLNDAAMGSECYNVSVFYVSHLFSWA 125  
 Db 71 LPHVEGLPPGAE-----TAELTPASAEILLKVALDLMOPOIKILLSHLKPHFVLFDAQEWL 127  
 Qy 126 AEVAREVDVPSALLWIEPATVFDVYFYFNGYADD-IDAGSDQIQLPNLQSLKQDLPSF 184  
 Db 128 PKMANGLGIKT-----VYVSVVALSTAFATLCPARVLEPKKVPSPLEDMKKPPL 175  
 Qy 185 LLP-SSPARFRLMKKEP-----DTLDKEPK--AKVLINTFDALETEQLKA 227  
 Db 176 GFPOTSVTSVTFEARDFLYFKSFHNGPTLYDRIQSLRGCSAILAKTCSQMEGPYIKY 235  
 Qy 228 ID---RYELISIGLIPSSIFSDGNDPSSSKSYGDLFRKADETYMDWLSKPSSEVVY 284  
 Db 236 VEAQFNKKPVFLIGVVP-----DPPSG-----KLEEKWATLANKFEGGTVIY 277  
 Qy 285 VSPGSLRLPKPQMBEIIAIGLSDTKSPVLWVIRNNEEGDEQAEAEKLLSFDRHGT 344  
 Db 278 CSFGSETFLTDQVKELALGLEQTGLPFLVL--NFPANVDVSAELNRLPGEF---LE 331  
 Qy 345 RL---GKIYT-WCSQDLVTHKSGVCGFVTHCGWNSAIESLACGVVPCPPQWFDQGTNAK 400  
 Db 332 RVKDGIIHSGWVQQQNTLAHSSVGYCHAGFSSVIEALVNDQCVWMLPKQGDILNAK 391  
 Qy 401 MIEDVWRSGVRVNEEGVVDREIKRCVSEVIKRE-----LRESAMWKG--LAKE 452  
 Db 392 LVSGDMEAGVEINRDEGDYFGKEDIKEAVEKVMVDVNDPFGKILREKQKWKPELLAND 451  
 Qy 453 AMDEERGSSMNL 465  
 Db 452 IQSYIGNLVNEM 464  
 -----  
 RESULT 13  
 UFOG\_MANES STANDARD; PRT; 394 AA.  
 ID UFOG\_MANES STANDARD; PRT; 394 AA.  
 AC Q40288;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 6 (EC 2.4.1.91) (UDP-GLUCOSE  
 DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 6) (FRAGMENT).



[illegible]

Qy	117	VYSHLFSWAEEAVREVDVP	SALLWIEPATVF-----DVYFYFNGVADDIDAGS	165
Db	116	V-GDAFYMTADAASAAGAPV	VAVTAASALLAHLRTDALRRDV-----GDQAASRA	166
Qy	166	DOIQLPN--LPOLSKDPLSFL	PSSPARPRTLMKEKFTLDKEPKA--KYLINTDEALE	221
Db	167	DELLVAHAGLGGYRVRLDP	GVVSGDFNVYISLLVHR--OAOQLPKAAATAVALNTE	224
Qy	222	TEQLKADRYELISIGLPISS	IFSDGNDPSSSNKSYGGDLPRKADETYMDLNSKPSS	281
Db	225	PPDLIAALAEPLNCLPLGP	YHLL--PGAETATDNAPAD-----PHGCLAWLDRRPAS	278
Qy	282	VVYVFSGLRLPKPQMEITA	IGLSOTKSPVLVYIRRNESGDGEQAEBEKKLLSFDRH	341
Db	279	VAVYFSGTNATARPDELQEL	AAGLEASGAPFLNSLR-----GVVAAAPRG	323
Qy	342	GTERL--CKIYTWCSOLDVLT	HKSVGGCFVTHCGHNSALESIACAGPVVCFPOHPDQGTNAK	400
Db	324	FLERAPGLVVPWAPQGVCL	RHAAGVAFVTHAGWASVWEGSSGVPMACRPFPGDQGTMNAR	383
Qy	401	MIEDVWRSGVRVRYNEEGG	VYVDRETKRCYSEVIKRE----LRESAMWKGLAKEAMDEE	457
Db	384	SVASVWGCFTAF-----D	GPMTGRCAVANAVATLLRGEDGERMAKQELQAMVYGKAFED	438
Qy	458	RGSSMNNLKNFI	469	
Db	439	-GGCRKNFDEFV	449	

Search completed: August 1, 2000, 21:32:20  
Job time: 6040 sec





```

Db      290  LTHGVLIATNRPFLWIVR-----EKNIPEKKKNRFLLEIRGSDR-GLVVWGCSTQAVL 340
Qy      360  THRKSGVCFYTHCGWNSAIESLACGVPVVCFFQWFQDGTNAKMIEDVWRSGYRVRVNNEGG 419
Db      341  AHCAGGCFYTHCGWNSTLESSEGGVPVAFQFADCTTAKLVEDTWRIQGVKKVGEED 400
Qy      420  VYDRREIKRCVSEVI-----KSELRESAMMKGLAKEMDDEERGSSWNNLNKFI 469
Db      401  -VDGEEIRRCLEKVMSSGGEAEEMRENAEKWAKMAVDAA-AEGGPSDLNLKGFV 452

RESULT  2
T02238
glucosyl transferase, jasmonate-induced - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 24-Nov-1999
C:Accession: T02238
R:Kojima, H.; Hashizume, K.; Imanishi, S.; Nakamura, K.
submitted to the EMBL Data Library, January 1997
A:Description: Jasmonate-induced potential glucosyltransferase from tobacco
A:Reference number: Z14633
A:Accession: T02238
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-467 <KOJ>
A:Cross-references: EMBL:AB000623; NID:d1094897; PID:d1019901
A:Experimental source: strain BY-2

```

RESULT 3  
 C71420  
 hypothetical protein - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 A:Variety: columbia  
 C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999  
 C:Accession: C71420  
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D  
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzensegger, T.; Pohl, T.M.; Terry, N.; G  
 avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B  
 Nature 391, 485-488, 1998  
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen  
 erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Paine, K.; Benes, V.; Rechman, S.;  
 C.; Chalkatzis, N.  
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t  
 A:Reference number: A71400; MUID:98121113  
 A:Accession: C71420  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-458 <BEV>  
 A:Cross-references: GB:297339; NID:g2244901; PID:g326935; PID:g2244911  
 C:Genetics:  
 A:Map position: 4COP9-4G3845  
 C:Superfamily: flavonol O3-glucosyltransferase  
  
 Query Match 36.0%; Score 895; DB 2; Length 458;  
 Best Local Similarity 42.0%; Pred. No. 1.3e-59;  
 Matches 204; Conservative 81; Mismatches 135; Indels 66; Gaps 17;

RESULT 4  
T00506  
indole-3-acetate beta-glucosyltransferase homolog T20D16.11 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999  
C:Accession: T00506

```

Query Match          25.7%; Score 639.5; DB 2; Length 438;
Best Local Similarity 35.3%; Pred. No. 1.8e-40;
Matches 165; Conservative 78; Mismatches 176; Indels 49; Gaps 16;

QY 15 QGHINPSLEFAKRLNTGYVDQVTFFTSVVALRMRRETD-PSSRIDFVAXXDSYDGLK 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 QGHLNPLMKFAKHLARTNL--HFTLATTEQARDLLSSTADEPHRPVOLA----FFSDGLP 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 74 KGD--DGKNYWMSEMRKGTAKLKDTLKLNDAAMGSCYNRVSVFYVSHLESWAAEVARE 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 KDDPRDPDTLAKSLKDGAKNLKIIBEKR-----FDCIISVPFT-----PWVPVAAA 109
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

110 HNICATLWQACAGFVYRYRYMKTNP -PDLEEDNQTVELPALPILVLRDPLSLMPS 168
189 SPARFRLMKKEFDTLDEKPKAKVLINTFOALEQOLKAI -DRYELISIGLPISSIFSD 247
169 QGANVNTLMAEFADCL -KDVK -WVLVNSFYELSEIETESMSDLKPIIPGLV -SPFL 224
248 GNDPSSNKSYGGDLFRADETYMWNLSKPESSVYVSGSLRLRKPQOMEIATIGSLD 307
225 GNDEKTLDMW - ----KVDDYCMELWDKQARSSWYISFGSILKSLSENQVETIATALKN 278
308 TKSPVLVIRNREGEDEQAEBEKLLSFDFRNGTRELKIVTWCSQLDVLTHKSVGCF 367
279 RGVFLVIRPKEGENVQVLOEMVK - ----EGKGVVTEWGOOEKTLSHMAISCF 328
368 VTHCGWNSATSLACGVVPVCFQWFDQGTNAKMIEDVWRSGVRVRYNEEGVVDREIK 427
329 ITHCGWNSTIETVYGVFPVAYPTWIDQDLARLLVDVFGIGVRKMKNDADIGELKVAEVE 388

```

Db 389 RCTEAVTEGPAADMRRRATELKHAAFSAM-SPGGSSAQNLDSFISDI 435

RESULT 6  
D71419  
probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1998  
C:Accession: D71419  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.;  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; G  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, E  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomene  
orheft, D.; Weeres, M.; Terzag, I.D.C.; Tschopp, T.; Palocz, Y.; Boeckmann, J.; Boeckmann, J.

C7 Charwazi33, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana  
A:Reference number: A71400; MUID:98121113  
A:Accession: D71419  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-484 <REV>  
A:Cross-references: GB:D97339; NID:g2244901; PID:e327472; PID:g2244905  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: Atcop9-3-glucosyltransferase

Query Match            25.7%   Score 637.5; DB 2; Length 484;  
Best Local Similarity 33.9%; Pred. No. 3e-40;  
Matches 169; Conservative 87; Mismatches 178; Indels 65; Gaps 19;

QY         3   NKRRHLLATPPAQGHINPSLEFAKRLNTGVVDQVTFETSVVALRMR-----FETDP 55  
| : : : | : ||| :|| : : : | : ||| : : ||| : : |||  
Db      10 NPIMVLVSFGQGHHVNPPLRLGLRIASKGLL--VTFTTTLWGKKRKQRANKIVDGE LKP 67

```

Qy 56 ----SSRIDFVAXXDSDGLKGDDGKN-----YMESEMRKRTKALKDTLIKLNDAAMGS 107
Db 68 VGSGRIRFEF-----FDEWAEADDORRADFSLYIAHLESVGIREYKLVRRYEAA-----117
Qy 108 ECYNRVSVFYVYSHLEWAAAEVAREVDVPAGALLWIEPATVDFVYFYFENG---YADDIDAG 164
Db 118 --NEPVSCLLNNFIPWCHVAEFPNIPCAVLWQSCACFSAYHYHQDSVSFPTEPE 175
Qy 165 SDQIQPLNLPOLSKQDLPSPFLPSSP-ARFRTLMKEKFDTLDOKEPAKAKYLINTFOALETE 223
Db 176 LD-VKLPVPLVKNDEIPSPHPSSRFTGFRQAILGQFNLSK--SFCVLIDSFDSLEQE 232
Qy 224 OLKAIDRAYELJ-----SIGPL--IPSSIIFSDGNDPSSSNKSYGGDLFRKADETYMDWLANSK 277
Db 233 ---VIDYMSLCPVKVTGPFVKVARTVSD-----VSGDICKSTDKC-LEWLDJR 278
Qy 278 PESVYVYVFGSLRLRPKPMOEIEAIGLSDTKSPVLWIRRNEDGEOBQAEKEKLSF 337
Db 279 PKSSVYISGTAYLAKQEQIEIEAHGVLSGLSFLWIR---PPPHDLKVETHVLQOE 334
Qy 338 FDRHGTRELKGIYVWCSQLDVLTHKSVGCFVTHHCGWNNSAIESLACGVVPVPCPQWDQGT 397
Db 335 LKESSAKGKGMIDWCQPQEQVLSSHPSVACFVTHCGWNSTMESLSSGVPVYVPCPQWDQVT 394
Qy 398 NAKMIEDVWRSGVRV--RVNEEGGVVDRRETKRCVSEVI--KSRELESAMMMKGLAKEA 453
Db 395 DAVYLDVFTGYVRLGCGATEERVVPEEVAEKLLEATVGERKABELRKNAKWKAEAE-BA 453
Qy 454 MDEERGSMNNLNKNFITRI 472
Db 454 AVAPGGSSDKNFRFEVEKL 472

```

RESULT 7  
E71419  
probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: columbiana  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999  
C:Accession: E71419  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkvist,  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terzaghi,  
A.; vanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.;  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Recl  
C.; Chludzinski, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arab  
A:Reference number: A71400; MUID:98121113  
A:Accession: E71419  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-479 <REV>  
A:Cross-references: GB:Z97339; MID:G2244901; PID:e326931; PID:G2244906  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: flavonol O3-glucosyltransferase

	Query Match	25.6%	Score 635;	DB 2;	Length 479;
	Best Local Similarity	32.6%;	Pred. No. 4.5e-40;		
	Matches 166;	Conservative 90;	Mismatches 173;	Indels 80;	Gaps 20;
Qy	6 HLLATPAGGHINPSLEFAKRLNTGYVDQVTFTSYAL-RMRPFETDSSRID----	60			
	: : :           : : :       : :     : :   :				
Db	8 HMVLSPGQGVHNP LLRLGLRIASKGLL--VTFTTEPKWGKKMR---QANKIQDGV L 61				
	: : :           : : :       : :     : :   :				
Qy	61 -----FVAXXDSYDDGLUKGDGGKNYNSEMKRGTKALKOTLIKLNDAAWGSECYN 111				
	: : :     : :   : :   : :   : :   : :   : :   :				
Db	62 KPVGLGTRTPEFFSDGFADDDKEKRFDFAFRPHLEAVGKQEIKN-LVKR-----YN 111				
	: : :     : :   : :   : :   : :   : :   : :   :				
Qy	112 R--VSFYVSHLSFWAAEVAREVDPVSALLWIEPATVDFDYFFYPNGYA-----DDID 162				
	: : : : :   : :   : :       : :   : :   : :   :				

Db	112	KEPVTCLINNAFVPCWCDVAEELHPSAVLWVSCACLTATAYYYYHHRLVKFPTKTEPDI - 170
Qy	163	AGSDIQIPLNLPQSKODLPFLPSSP-ARFRTLAKKEFDTLDKPKAKVLINTYDALE 221
Db	171	---SVEIPLCLPKLKHDEISFLHPSSPYTAFGDIILDLQKRFENHKSFLYFDITFRELE 226
Qy	222	---TEQLKAIIDRYBLIS-IGPLIPSSIFSDGNPSSNKSYGGLFRKADETYMDWLNSK 277
Db	227	KDIMHNSQLCPQAIISPVGFL-----FKMAQTLSSDVK---GDISEPASDC-NEWLDSR 277
Qy	278	PSSVYVYSGSLRLRLPKPQMEETATGLSDTKSPVLVWIRNNEEGDEQEQAEERKLLSF 337
Db	278	EPSSVYVYSGFTIANLKOBEETAHGVLSSGLSVLWVRPPMEG-----TF 324
Qy	338	FDRE---GTERLGKIYVTCSDQLDLVTHKSVGCFVYTHCGWNSAIESLACQVPPVCPQPF 393
Db	325	VEPHVLPRELEEGKIYVENCPOERVLAHPALACFLHCGWNSWEALTAGVPPVCPQWG 394
Qy	394	DOGTNAKMIEDVWRSGYRV-RVNEEGGVDRREIKRCVSEVI---KSRELRESAMWKG 449
Db	385	DQVTDVAVLADVFTGVRGLGRGAAEEMIVSREVVAEKLLEATVGEKAVELRENARWKA 444
Qy	450	AKEMDEERGSSMNLNKFTRIINENAS 478
Db	445	A-EAAVADGGSSDMNFKEFYDKLWTKHVT 472
RESULT	8	
A54739		Indole-3-acetate beta-glucosyltransferase (EC 2.4.1.121) - maize
C:Species:		zea mays (maize)
C:Date:		28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 29-Oct-1999
C:Accession:		A54739
R:Szerszen, J.B.;		Szczylowski, K.; Bandurski, R.S.
Science		265, 1699-1701, 1994
A:Title:		taglu, a gene from Zea mays involved in conjugation of growth hormone indole
A:Reference number:		A54739; PMID:94367368
A:Accession:		A54739
A:Status:		preliminary; not compared with conceptual translation
A:Molecule type:		mRNA
A:Residues:		1-471 <SIZE>
A:Cross-references:		GB:L34847; NID:g548194; PIDN:AAAS9054.1; PID:g548195
C:Superfamily:		flavonol O3-glucosyltransferase
C:Keywords:		glycosyltransferase; hexosyltransferase
Query Match	25.4%	Score 631; DB 2; Length 471;
Best Local Similarity	33.1%	Pred. No. 8.7e-40;
Matches	163; Conservative	78; Mismatches 197; Indels 46; Gaps 16;
Qy	6	HILLATPAGCHINPSLEFAKRLINTGYVDQVFTTSVVALRRMRPETDPSRIDFVAXX 65
Db	4	HVLVPPPGGHHMPPQFAKRLASKVA---TTLVTRFQRTADVADHPAM---VEALS 58
Qy	66	DSYDDGLKGDGKNYSEMRKRGTRAKLTKLINDAAMGSECYNRVSPVYSHLFSWA 125
Db	59	DGHDEG---GFASAAGVAEYLEKQAAAAASLASLVEARASSA--DAFTCVVYDSYEDWV 113
Qy	126	AEVAREVDVPSALLWIEPATVFDVYYPYFNG-----YADDIDAGSDQIQLP----NL 173
Db	114	LPVARRMGLPAVPPESTQSCAVSAVYHFSGRLAVPPGAAADSGDGAGAAALSEAFGL 173
Qy	174	POLSKDLPFLPSSPAPRTLMKEFDTLDKPEKAK-VLINTFDALTEQLKAIIDRY- 231
Db	174	PEMERSELPSEFVDHGP--YPTAMQAIKGFUAGKDDWVLFNFSFELETEVLGLTKYL 231
Qy	232	ELISGLIPSSIFSDGNPSSNNK-SYGDLPFRKADETYMDLNSKPSVYVYSGSL 290
Db	232	KARAIGFCVP--LPTAGRTAGNCRITYGANLV-KPEDACTKWLDTKPDRESVAYVSGSL 288
Qy	291	LRLPKPQMEETATGLSDTKSPVLVWIRNNEEGDEQEQAEERKLLSFFDRHGTGR-LGKI 349
Db	289	ASLGNQAEKELRGLLAAGKPFLLWVR-----ASDEHQVPRYLLAEATATGAAMY 338

Qy	350	VTWCSQLDVLTHKSVGCFYTHCGWNSAISLSLACGVPVVCFFQWFOGQGNAKMIEDVWRSG	409
Db	339	VPMCPQLDVLHAHPAVGCFYTHCGWNSLTLEALSGFVPMVAMALWTQDPTNARNVELAWGAG	398
Qy	410	VYRVNVEGGVDRRRIKKVCSEVSKSR- ---LRESAMMMKGLAKEAMDEERGSSMNNL	465
Db	399	VRARRDAGAGVFLRGEVERCVRAVMDGGGAASAARKAKAGENWRDRAAV-APGSSDRNL	457
Qy	466	KNFI 469	
Db	458	DEFV 461	
RESULT	9		
T00639			
		hypothetical protein F316.2 - Arabidopsis thaliana	
		C:Species: Arabidopsis thaliana (mouse-ear cress)	
		C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999	
		C:Accession: T00639	
		R:Pederspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.;	
		Ysotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.	
		submitted to the EMBL data Library, February 1998	
		A:Reference number: Z14197	
		A:Accession: T00639	
		A:Status: translated from GB/EMBL/DDBJ	
		A:Molecule type: DNA	
		A:Residues: 1-460 <PED>	
		A:Cross-references: EMBL:AC002396; NID:g2829918; PID:g2829862; GSPDB:GN0005	
		C:Genetics:	
		A:Gene: ATSP.F316.2	
		A:Map position: 1	
		A:Introns: 219/3	
		C:Superfamily: flavonol O3-glucosyltransferase	

[illegible]

```

Db 450 SINEFI 455
:: II
RESULT 10
T00511
Indole-3-acetate beta-glucosyltransferase homolog T20D16.16 - Arabidopsis thaliana (mouse-ear cress)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00511
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, J.L.
submitted to the EMBL Data Library, November 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence
A:Reference number: Z14159
A:Accession: T00511
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-453 <R0UD>
A:Cross-references: EMBL:AC002391; NID:g2642427; PID:g2642442
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 286/2
A:Note: T20D16.16
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 23.9%; Score 593.5; DB 2; Length 453;
Best Local Similarity 33.2%; Pred. No. 5.4e-37;
Matches 159; Conservative 80; Mismatches 181; Indels 59; Gaps

```

[illegible]

RESULT 11  
F71419  
probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999  
C:Accession: F71419  
P:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkvist,  
R.; Weidler, H.; Weidler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terras,  
D.

avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans  
C.; Chaiwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A:Reference number: A71400; MUID:98121113  
A:Accession: F71419  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-475 <BEV>  
A:Cross-references: GB:297339; NID:g2244901; PID:e326932; PID:g2244907  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 23.8%; Score 590.5; DB 2; Length 475;  
Best Local Similarity 32.5%; Pred. No. 9.7e-37;  
Matches 162; Conservative 84; Mismatches 180; Indels 73; Gaps 19;

QY 6 HILLATPAOCHINPSLEFAKRLNTGYVDVFTTSYVAL-RMRFTDPSS----- 57  
DB 9 HMLVSPFGQGHISPLRLGLIIASKGLI--VTFTTEEPGLKMRQANNIQQGVLPVG 66  
QY 58 ----RIDVAXXDYDDGLKKGDDGKNYMSMRKRGTKALKDTLIKLNDAAMGSCYKNRV 113  
DB 67 LGFLRFEF-----PEDGFVYKEDFDLLQKLSLEVSKRKINLVKKYKQPV--RCLINN 118  
QY 114 SFVYVSHLFAAEVAREVDVPSALLWIEPATVFDVYFYNGYADDIDAGSDQI--QLP 171  
DB 119 AFV-----PWVCIAEQLQIPSAVLWVQSCAAIAAYVYHHQLVKFTEPEITVDVP 172  
QY 172 NLPQSKOD-LPSFLLPSP-ARFTLMKEKFDTLKPEKAKVLINTFADALETEOL---- 225  
DB 173 FKPTLKHDEIPSFHPSPSSIGTGTLQIKRLHK--PFSVLITETFOELEKDTIDHMS 230  
QY 226 KAIRYELISIGPL--IPSSIFDGNDDPSSNKSXGGLDFRKADTYMDLNSKPSVW 283  
DB 231 QLCQVNPNGPLFTMAKTRSD-----IKGDI-SKPDSDCIWLDSDREPSVW 279  
QY 284 YVSGSLRLPKPQMEETAIGLSDTKSPVLWVIRNNEGDEQE-----QAEEEKILLSFFD 339  
DB 280 YISFTLAFLKQNDIETAHGLNSGLSLVLRPLGLELAIEPHVLPLEEEK----- 333  
QY 340 RHGTERLGIKIVTWSQLDVLTKSGCCVTHCGWNSAIESLACGVPVVCPPQWFDQGTNA 399  
DB 334 -----GKIVEMCQEKVLAHPAVACFLSHCGWNSMTHEALTSGVPVTCFQWGDQVTNA 386  
QY 400 KMIEDVRSRGRV-RVNEEGGVVDRREIKRCVSEVI---KSRLERESAMMKGLAKEAMD 455  
DB 387 VYMDVFTGLRLSGASDERIVPREVAERLLEATVGEKAVELRENARRWKEAESAV- 445  
QY 456 BERSSMMNLNFTIRIN 474  
DB 446 AYGGTSENFQEFVKLVD 464

RESULT 12

T00584  
Indole-3-acetate beta-glucosyltransferase homolog T27E13.12 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999

C:Accession: T00584

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul

submitted to the EMBL Data Library, May 1998

A:Description: Arabidopsis thaliana chromosome II BAC T27E13 genomic sequence.

A:Reference number: Z14177

A:Accession: T00584

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-455 <ROU>

A:Cross-references: EMBL:AC004165; NID:g3150396; PID:g3150406

A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 2  
A:Introns: 166/1  
A:Note: T27E13.12  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 20.1%; Score 499; DB 2; Length 455;  
Best Local Similarity 31.2%; Pred. No. 6.8e-30;  
Matches 153; Conservative 74; Mismatches 198; Indels 66; Gaps 17;

QY 5 RHILLATPAOCHINPSLEFAKRLNTGYVDVFTTSYVALRRMRFTDPSSRIDFVAX 64  
DB 12 RHVAMPYPGRGHINPMNLCRLVRRYPNLHVTFTVEELGFGPDKP-KDRHFSTL 70  
QY 65 XDSYDDGLKKGDDGKNYMSMRKRGTKALKDTLIKLNDAAMGSECYNRSVYVYSHLF-S 123  
DB 71 PNLIPSELVRAKDFIGFIDAVYTRLEPPEKLLDLSNPPPS-----VIFADTVVI 121  
QY 124 WAAEAREVDVPSALLWIEPATVFDVY---YFYNGYADDIDAGSDQI--QLPQLS 177  
DB 122 WAVRVGRKRNPVVSMTMSATILSFLLHSDLLISHGHA-LFEPSEEEVDVYVPGLSPTK 180  
QY 178 KODLPSFLLPSPARFTLMKEKFDTLKPEKAKVLINTFADALETEOLKAIDRY----- 231  
DB 181 LRDLPIFDGSDRVFKT-AKLCFDEL---PGARSLLFT-TAYELEH-KAIDFTSKLDI 234  
QY 232 ELISIGPLIPSSIFDGNDDPSSNKSXGGLDFRKADTYMDLNSKPSVYVYSGSL 291  
DB 235 PYVATGLPIPFELSQVNDKPN-----YIQWLERQPEGSLVYISQGSPL 280  
QY 292 RLPKPQMEETAIGLSDTKSPVLWVIRNNEGDEQEAEKEELLSFFDRHGTERLGIKIVT 351  
DB 281 SVSEAQMEIEVKGRLRESGVRLWVAR---GGELKLEALE-----GSLGVVYS 325  
QY 352 WCSQLDVLTKHSGCVFTHCGWNSAIESLACGVPVVCPPQWFDQGTNAKMIEDVRSRGRV 411  
DB 326 WCDQLRVLCRAVGGFWTHCGFNSLGLYSGVPLAPLFWDQILNAKMIYEDVRVGR 385  
QY 412 V-RVNEEGGVVDRREIKRCVS-----EVIKSRELRESAMMKGLAKEAMDEERGSMNML 465  
DB 386 IERTKKNELLIGREBIKEVVKRFMDRESEGEKEMRRACDLSEISRGAV-AKSGSSNVNI 444  
QY 466 KNFIRIINEN 476  
DB 445 DEFVRHITNTN 455

RESULT 13

T45603

glucosyltransferase-like protein - Arabidopsis thaliana

N:Alternate names: protein F12A12.180

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000

C:Accession: T45603

R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.;

submitted to the Protein Sequence Database, December 1999

A:Reference number: Z23008

A:Accession: T45603

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-453 <CHO>

A:Cross-references: EMBL:AL133314

A:Experimental source: cultivar Columbia; BAC clone F12A12

C:Genetics:

A:Map position: 3

A:Introns: 161/1

A:Note: F12A12.180

C:Superfamily: flavonol O3-glucosyltransferase

Query Match 19.8%; Score 491.5; DB 2; Length 453;







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 14:38:35 ; Search time 3616.02 Seconds  
(without alignments)  
1752.458 Million cell updates/sec

Title: US-09-147-955-9  
Perfect score: 1437  
Sequence: 1 ttcaaaactcataacgtgat.....tttatttgagtaaaaaaaaa 1437

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues  
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*  
1: em\_est1.\*  
2: em\_est2.\*  
3: em\_est3.\*  
4: em\_est4.\*  
5: em\_est5.\*  
6: em\_est6.\*  
7: em\_est7.\*  
8: em\_est8.\*  
9: em\_est9.\*  
10: em\_est10.\*  
11: em\_est11.\*  
12: em\_est12.\*  
13: em\_est13.\*  
14: em\_est14.\*  
15: em\_est15.\*  
16: em\_est16.\*  
17: em\_est17.\*  
18: em\_est18.\*  
19: em\_est19.\*  
20: gb\_est1.\*  
21: gb\_est2.\*  
22: gb\_est3.\*  
23: gb\_est4.\*  
24: gb\_est5.\*  
25: gb\_est6.\*  
26: gb\_est7.\*  
27: gb\_est8.\*  
28: gb\_est9.\*  
29: gb\_est10.\*  
30: gb\_est11.\*  
31: gb\_est12.\*  
32: gb\_est13.\*  
33: gb\_est14.\*  
34: gb\_est15.\*  
35: gb\_est16.\*  
36: gb\_est17.\*  
37: gb\_est18.\*  
38: gb\_est19.\*  
39: gb\_est20.\*  
40: gb\_est21.\*  
41: gb\_est22.\*  
42: gb\_est23.\*  
43: gb\_est24.\*  
44: gb\_est25.\*  
45: gb\_est26.\*  
46: gb\_est27.\*  
47: gb\_est28.\*  
48: gb\_est29.\*  
49: gb\_est30.\*  
50: gb\_est31.\*  
51: gb\_est32.\*  
52: em\_est20.\*  
53: em\_est21.\*  
54: em\_est22.\*  
55: em\_est23.\*  
56: em\_est24.\*  
57: em\_est25.\*  
58: em\_est26.\*  
59: gb\_est33.\*  
60: gb\_est34.\*  
61: gb\_est35.\*  
62: gb\_est36.\*  
63: gb\_est37.\*  
64: gb\_est38.\*  
65: em\_est27.\*  
66: em\_est28.\*  
67: em\_est29.\*  
68: em\_est30.\*  
69: gb\_est39.\*  
70: gb\_est40.\*  
71: gb\_est41.\*  
72: gb\_est42.\*  
73: gb\_est43.\*  
74: gb\_est44.\*  
75: em\_est31.\*  
76: em\_est32.\*  
77: em\_est33.\*  
78: em\_est34.\*  
79: gb\_est45.\*  
80: gb\_est46.\*  
81: gb\_est47.\*  
82: em\_est35.\*  
83: em\_est36.\*  
84: em\_est37.\*  
85: gb\_est48.\*  
86: gb\_est49.\*  
87: gb\_est50.\*  
88: gb\_est51.\*  
89: gb\_est52.\*  
90: gb\_est53.\*  
91: gb\_est54.\*  
92: gb\_est55.\*  
93: gb\_gss1.\*  
94: gb\_gss2.\*  
95: gb\_gss3.\*  
96: gb\_gss4.\*  
97: em\_gss1.\*  
98: em\_gss2.\*  
99: em\_gss3.\*  
100: em\_gss4.\*  
101: gb\_gss5.\*  
102: gb\_gss6.\*  
103: gb\_gss7.\*  
104: gb\_gss8.\*  
105: gb\_gss9.\*  
106: em\_gss5.\*  
107: em\_gss6.\*  
108: em\_gss7.\*  
109: em\_gss8.\*  
110: em\_gss9.\*  
111: em\_gss10.\*  
112: em\_gss11.\*  
113: gb\_gss10.\*  
114: gb\_gss11.\*  
115: em\_gss12.\*  
116: gb\_gss12.\*



```

Qy 941 gagaaatgccctctcaatcaatgcagtgctctactggacttttggcagcctatgtttgta 1000
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 504 GCGAATGAGTTGAAGTCCATTCACTGCTAGCTGGAGCAGCACTTCCTCTTGTGGATT 563
Qy 1001 tgcactgtccagctctctctcttcacacacaccccttttctctcaagatgagtgccg 1060
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 564 TGCATTGGCGATGTTCTCTTCTACTCATATTGTTCCCAATTTTGGCTGAAGAGGATGGTGC 623
Qy 1061 tgcattttcaatttccactgacatgatatgctgggctgttgcaataggacatt 1120
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 624 TACAATGTAATCTAGCCTTGCCTTACCTCTGATATGGGGCTGTGTAATTCGTATCTT 683
Qy 1121 catatacaaccaggaggttgattgtttactattt 1156
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 684 TGTATACCATCAGAAGGTTGACTGGATATATTGCT 719

RESULT 2
LOCUS C72454 459 bp mRNA EST 22-SEP-1997
DEFINITION C72454 Rice panicle at flowering stage Oryza sativa cDNA clone
E1644_1A, mRNA sequence.
ACCESSION C72454
VERSION C72454.1 GI:2427991
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Mangoliophyta; Liliopsida; Poales; Poaceae; Oryza.
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from panicle at flowering stage
JOURNAL Unpublished (1996)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp.

FEATURES
    source
        1..459
            /organism="Oryza sativa"
            /strain="Nipponbare"
            /db_xref="taxon:4530"
            /clone="E1644_1A"
            /note="Rice panicle at flowering stage"
            /dev_stage="flowering stage"
            /note="Organ: panicle; Rice cDNA from panicle at flowering
stage"

BASE COUNT 110 a 88 c 110 g 146 t 5 others
ORIGIN

Query Match 7.18; Score 101.4; DB 81; Length 459;
Best Local Similarity 57.18; Pred. No. 3.5e-17;
Matches 205; Conservative 0; Mismatches 151; Indels 3; Gaps 1;

Qy 462 tacactgttggccctctctatggaccactctgtcttaccgcccgcacaaattcttg 521
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 8 TACATCTTCTCTCCCTTATCTAGGAGGTATCTCATCTACCGGAGACGCCCTCAGC 67
Qy 522 gtccatgtgactggtactgctctggtgggttcgtggagcgtccacggaattactctgtt 581
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 68 ATAAATGTACTACTATTTCCTTGGGATTTATGACGTGGAAGCAATATATAGTT 127
Qy 582 aataaagcattcgatgacatcgattacgagtgtagcatactggttggtgacaatc 641
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128 GTGAAGTCTTATCAGTACACATCTTTGACAAGCGTGATGCTCTTGGATTGCTGTCMAAT 187
Qy 642 gtgtgggtccatctcttacctggtatgttctctaggccacaaaatactctgtataccagttt 701

```

```

Db 188 CCTGTGTGATTATTCTTACATGGATATTTTGAAGACCAAGTATGGCTCAGGAAGTTC 247
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 702 gtccgtgctgctatttgcgtaggagccctcctcctcgtcgtctcttcctcagactcaggggtc 761
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 248 ATCGTGTGTGGATTGCGTGGCTGAGCTTATATTAGTAGTATTTTTCAGATGTCATGCC 307
Qy 762 actcct---gcgtgttgaatcctcttgggtgatttcttcttcataacaggctctat 817
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 308 TCTGATCAGCTAGACCCACCCCTTGAAGGAGAGATTTGCTCGTAATCTTTGGGTCCAT 366
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 3
LOCUS A1973593 604 bp mRNA EST 18-APR-2000
DEFINITION sc88h08.y1 Gm-cl019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl019-472 5' similar to TR:059785 059785 HYPOTHETICAL 56.8 KD
PROTEIN. ;, mRNA sequence.
ACCESSION A1973593
VERSION A1973593.1 GI:5770419
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Mangoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 604)
AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Erpelding,J., Coryell,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
TITLE On Aug 21, 1998 this sequence version replaced gi:3704615.
JOURNAL Contact: Shoemaker R/Public Soybean EST Project
COMMENT Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Insert Length: 1451 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 436.
FEATURES
    source
        1..604
            /organism="Glycine max"
            /db_xref="taxon:3847"
            /clone="GENOME SYSTEMS CLONE ID: Gm-cl019-472"
            /clone_lib="Gm-cl019"
            /tissue_type="immature seed coats of greenhouse grown
plants"
            /lab_host="DH10B (Gibco BRL)"
            /note="Vector: pSPORT1 (Life Technologies); Site_1: Not I;
Site_2: Sal I; This cDNA library was constructed from mRNA
isolated from immature seed coats (200-300 mgs) of
greenhouse grown plants. The library was prepared using
the Life Technologies pSuperScript cDNA library
construction kit. Complementary DNA was synthesized from
mRNA using a poly (dT) sequence with a Not I restriction
site. Sal I linker adapters were ligated to the
blunt-ended cDNA fragments followed by Not I digestion.
The cDNA fragments were directionally cloned into the Not
I-Sal I restriction site of the pSPORT1 vector. The
ligated cDNA fragments were transformed into E.coli
Electromax DH10B host cells (Gibco BRL). This library was

```





Contact: Blaxter ML  
Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
3JT, UK.  
Tel: +44 131 650 6760  
Tel: +44 131 670 5450  
Fax: +44 131 670 5450  
Email: mark.blaxter@ed.ac.uk  
The ABI trace of this sequence can be viewed at  
<http://www.sanger.ac.uk/brugia/AFC/MBAFWC8D08T3.html>  
Seq primer: T3.

```

FEATURES
source
location/Qualifiers
1. .409
/organism="Brugia malayi"
/db_xref="taxon:6279"
/clone="AFCW8D08"
/clone_lib="Brugia malayi adult female cDNA
(SAW96MLW-BmAF)"
/sex="female"
/dev_stage="adult"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site.1: ECO RI; Site.2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from approximately 50 adult females
isolated from the peritoneal cavity of jirds and
converted to double-stranded cDNA using reverse
transcriptase and oligo(dT) followed by RNase H and DNA
pol I. The library has 5 x 10E6 independent recombinants
and the average insert size is ~900bp. The library was
constructed by Michelle Lizotte-Waniewski. The
library is available from Dr.S.A.Williams, email:

```

```

BASE COUNT      104 a      60 c      86 g      159 t
ORIGIN
genome.BS.Matched

Query Match          4.9%; Score 70.2; DB 31; Length 409;
Best Local Similarity 50.4%; Pred. No. 1.2e-08;
Matches 205; Conservative 0; Mismatches 193; Indels 9; Gaps 1;

QY 526 ggattgttgacaatcgtggtcccatcatctttacatggatgttctctaggcacaaaata 685
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GGATTTGCTCAACGATACCGATGGTATTACTATTATCGTGGTATTCCCTTTACACGCCTA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 686 ctctgtataccagtttgtcggtgcctctatttgttaggagggcctccttcctcg|gct--- 742
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 CCTCTTGACACATAATAATTGGGTGTGGAATATGCTTGTTCGGAATTACCGTTC|TATTTG 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 743 -----tctttccagctcaaggggtaactgctgctcgtggttcgaatacctcttttgggtgatatt 796
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GGCTGATGCTTTTGGAAAGGAAGGGTACTTCTGGTGATTTCTTAATCGAGTATTAGSTGATGT 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 797 tcctgtcataaacaggctctattttgttccacctcagcactgttggtcaggaaatactgcgt 856
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 ATTAGTTTGATTGGTTCAGTTC|TATATGCAATTTGGTAANTCTTAGTGAGGAATTTTGGT 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 857 gaagagagaagatcgttatattaagtagtagcaaatgatcogggtgatttggdtagtctcaatcag 916
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 GAAACAAAATAGTCGAATTTGAATATCTTTGGTATGGTTGGCTTATTTTGGCTCAATCATCTC 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 917 tgcaacsagattactgtcctggagagaaatgcctctcatcaatgcagtgggtctactgg 976
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TGGTATACAACCTCGTACTTTAGAGCATCTGCAATTAGCATCAATTAATTGGTCAGGAAT 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 977 acttttggcagcctatgttggtttatgcactgtcccagcttctctctct 1023
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GATCATAAATATAC|TACCCTTTTATTTTGGTGCATGTATGTTCTTATTTCT 407
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT	8
LOCUS	AU066704
LENGTH	960 bp
DEFINITION	Suqano mouse brain mncb Mus musculus EST
VERSION	02-JUN-1999
CLONE	clone MNCB-1380

5', mRNA sequence.	
ACCESSION	AU066704
VERSION	AU066704.1 GI:4967441
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	house musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Hashimoto, K., Kusuda, J., Toyoda, A., Tanuma, R., Ito, A., Hirata, M., Suzuki, Y., Sasaki, M. and Sugano, S.
TITLE	Isolation of full-length cDNA clones from a mouse brain cDNA library made by oligo-capping method
JOURNAL	Unpublished (1999)
COMMENT	On Apr 7, 1998 this sequence version replaced gi.3036257. Contact: Katsuyuki Hashimoto Division of Genetic Resources National Institute of Infectious Diseases 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan Email: khashienh.go.jp URL: <a href="http://www.nih.go.jp/yoken/genbank/">http://www.nih.go.jp/yoken/genbank/</a> .

FEATURES	SOURCE
----------	--------

```

/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="MNCB-1380"
/clone_lib="Sugano mouse brain mncb"
/sex="female"
/dev_stage="adult"
/lab_host="TOP10"
/notes="Organ: Brain; Vector: pME18S-FL3; 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTCGCCCTTTTTTTTTTTT] double-stranded cDNA was
ligated to a DraIII adaptor [TGTTCGCCCTACTGG] , digested
and cloned into distinct DraIII sites of the pME18S-FL3.
XhoI sites just outside the DraIII sites can be used to
isolate the cDNA insert. Size selection was performed to
exclude fragments <1.5 kb. Library was constructed by
Sugano et al. (University of Tokyo, Institute of Medical
Science). Custom primer for sequencing: 5' end primer
[CTTCTGCCTCTAAAGCTGG]"

```

BASE COUNT	250 a	215 c	214 g	230 t	51 others
ORIGIN					

Query Match 4.5%; Score 64.4; DB 48; Length 950;  
Best Local Similarity 48.6%; Pred. No. 5.7e-07;

[illegible]



```

RESULT 11
W41743      1001 bp      mRNA      EST      20-MAY-1996
LOCUS      mc63a08.r1 Soares mouse embryo NbMEL3.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:353174 5', mRNA sequence.
ACCESSION  W41743
VERSION     W41743.1 GI:1325772
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 1001)
AUTHORS    Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,
            Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
            Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,
            Theising B., Wylie T., Lennon G., Soares B., Wilson R. and
            Waterston R.
            The WashU-HHMI Mouse EST Project
            Unpublished (1996)
            On Apr 14, 1993 this sequence version replaced gi:635814.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:224974
            Seq primer: EntPrimer
            High quality sequence stop: 409.
FEATURES   source
            Location/Qualifiers
            1..1001
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="IMAGE:353174"
            /clone_lib="Soares mouse embryo NbMEL3.5 14.5"
            /sex="unknown"
            /tissue_type="embryo"
            /dev_stage="13.5-14.5dpc total fetus"
            /lab_host="DH10B"
            /note="Vector: pT7p3D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
            was primed with a Not I - oligo(dT) primer [5'
            TGTACCAATCTGAAGTGGGAGCGCGGAAATTTTTTTTTTTTTTTTTT
            T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
            14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
            State Univ. from 2 ]; double-stranded cDNA was ligated to
            Eco RI adaptors (Pharmacia), digested with Not I and
            cloned into the Not I and Eco RI sites of the modified
            pT7T3 vector. Library went through one round of
            normalization, and was constructed by Bento Soares and
            M.Fatima Bonaldo."
BASE COUNT 202 a 200 c 226 g 265 t 108 others
ORIGIN
Query Match 3.9%; Score 55.4; DB 91; Length 1001;
Best Local Similarity 47.4%; Pred. No. 0.00017;
Matches 207; Conservative 0; Mismatches 221; Indels 9; Gaps 1;

QY 600 acatcgattacgagtgtagcactggtgttgagacaatcgtgtggtccatcatttt 659
|| || || || || || || || || || || || || || || || || || || ||
Db 1 ACGACTCTACCAAGTGTCCAGACTCTTGGATTGCTTTGGGATTCCTGTGTGATGGCTCTC 60

QY 660 acatgagtgcttcagcacaataactcgtgtataccagtttctgctgctatttgt 719
|| || || || || || || || || || || || || || || || || || || ||
Db 61 TCCTGTGTTATTCTCCGGGAGACATACAAAGTATCCATTCATTCCTGTGTGTGCTGT 120

```

---

```

QY 720 gtagggagccctcctcgtgcttcttccgactca-----gggggctactgctgct 770
Db 121 CTCTGGGTGTAGGGACCATGGTTGGTGACACATATTAGTGGCGAGAGATAATTCA 180

QY 771 ggtcgaatcctcttttggggtgatttcttgctcacaacaggtcctattttgttcacactc 830
|| || || || || || || || || || || || || || || || || || || ||
Db 181 GGGAGTGTATGTTCTTATTGGTGACATCTTAGTCTCTGCTTGGGCTTCCCTGTATGCTGTG 240

QY 831 agcactgttggtcaggaataactcgtggaagagaagatcgattatgaagttagcaatg 890
|| || || || || || || || || || || || || || || || || || || ||
Db 241 TCTAAGCTGTGTGAAGATACATCGTGAAGAGCTGAGCAGACAGGAATTTTATAGGAATG 300

QY 891 atcggtgtattgtatgctcatcctcagtgcaacgagattactgtgctggagagaaagcc 950
|| || || || || || || || || || || || || || || || || || || ||
Db 301 GTGGGACTATTGGAAACAATCATCAGCGGCATACAGCTATTGATTGGGAATAAAGAC 360

QY 951 ctctcatcaatgacgtggtctactgacatttggcgacctatgttttattgacactgtcc 1010
|| || || || || || || || || || || || || || || || || || || ||
Db 361 ATTGCCCGGATTTCAGTGGGACTGGAATAATTGCCCTGTATTATTGGCATTTGCCCTCTGT 420

QY 1011 agcttctctcttctgcac 1027
Db 421 ATGTTTGGCTGTACAC 437

RESULT 12
W70677      488 bp      mRNA      EST      17-JUN-1996
LOCUS      me42c04.r1 Soares mouse embryo NbMEL3.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:390150 5', mRNA sequence.
ACCESSION  W70677
VERSION     W70677.1 GI:1380829
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 488)
AUTHORS    Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,
            Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
            Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,
            Theising B., Wylie T., Lennon G., Soares B., Wilson R. and
            Waterston R.
            The WashU-HHMI Mouse EST Project
            Unpublished (1996)
            On Sep 12, 1996 this sequence version replaced gi:1290773.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:241982
            Seq primer: EntPrimer
            High quality sequence stop: 340.
FEATURES   source
            Location/Qualifiers
            1..488
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="IMAGE:390150"
            /clone_lib="Soares mouse embryo NbMEL3.5 14.5"
            /sex="unknown"
            /tissue_type="embryo"
            /dev_stage="13.5-14.5dpc total fetus"
            /lab_host="DH10B"
            /note="Vector: pT7p3D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
            was primed with a Not I - oligo(dT) primer [5'
            TGTACCAATCTGAAGTGGGAGCGCGGAAATTTTTTTTTTTTTTTTTT
            T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
            14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
            State Univ. from 2 ]; double-stranded cDNA was ligated to
            Eco RI adaptors (Pharmacia), digested with Not I and
            cloned into the Not I and Eco RI sites of the modified
            pT7T3 vector. Library went through one round of
            normalization, and was constructed by Bento Soares and
            M.Fatima Bonaldo."

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:18:36 ; Search time 152.01 seconds  
(without alignments)  
1300.360 Million cell updates/sec

Title: US-09-147-955-9

Perfect score: 1437

Sequence: 1 ttcaaaactcataacgtgat.....tttatttgagtaaaaaaaa 1437

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/PCTRUS\_COMB.seq:\*
- 7: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	3.1	7218	1	US-08-232-463-14
C 2	43.6	3.0	7218	1	US-08-232-463-14
C 3	35.8	2.5	480	1	US-08-282-581-4
C 4	35.8	2.5	480	1	US-08-550-544-4
C 5	34.8	2.4	4951	4	US-08-867-030B-5
C 6	34.8	2.4	4951	6	PCT-US95-06119-5
C 7	33	2.3	1297	1	US-08-370-975B-12
C 8	33	2.3	26764	1	US-08-370-975B-1
C 9	32.8	2.3	19124	4	US-08-487-826B-13
C 10	32.4	2.3	11225	7	5182210-9
C 11	32.2	2.2	2472	5	US-08-335-844A-7
C 12	32.2	2.2	3358	5	US-08-335-844A-20
C 13	32	2.2	729	7	5182210-21
C 14	32	2.2	5138	3	US-08-476-062A-39
C 15	31.2	2.2	2080	1	US-08-077-935-18
C 16	31.2	2.2	2080	1	US-08-461-599-18
C 17	31.2	2.2	2080	1	US-08-461-621-18
C 18	31.2	2.2	2080	2	US-08-465-334-18
C 19	31	2.2	1298	2	US-08-473-157A-2
C 20	31	2.2	10079	4	US-08-476-866-20
C 21	30.8	2.1	179	5	US-09-065-474-100
C 22	30.8	2.1	179	5	US-09-065-474-101
C 23	30.8	2.1	810	2	US-09-031-485-9
C 24	30.8	2.1	810	2	US-09-031-485-10
C 25	30.8	2.1	810	2	US-08-847-429A-9
C 26	30.8	2.1	810	2	US-08-847-429A-10

27	30.8	2.1	810	5	US-09-065-474-9	Sequence 9, Appli
C 28	30.8	2.1	810	5	US-09-065-474-10	Sequence 10, Appl
C 29	30.8	2.1	864	5	US-09-065-474-144	Sequence 144, App
C 30	30.8	2.1	864	5	US-09-065-474-146	Sequence 146, App
C 31	30.8	2.1	936	2	US-09-031-485-4	Sequence 4, Appli
C 32	30.8	2.1	936	2	US-09-031-485-5	Sequence 5, Appli
C 33	30.8	2.1	936	2	US-08-847-429A-4	Sequence 4, Appli
C 34	30.8	2.1	936	2	US-08-847-429A-5	Sequence 5, Appli
C 35	30.8	2.1	936	5	US-09-065-474-4	Sequence 4, Appli
C 36	30.8	2.1	936	5	US-09-065-474-5	Sequence 5, Appli
C 37	30.8	2.1	937	2	US-09-031-485-1	Sequence 1, Appli
C 38	30.8	2.1	937	2	US-09-031-485-3	Sequence 3, Appli
C 39	30.8	2.1	937	2	US-08-847-429A-1	Sequence 1, Appli
C 40	30.8	2.1	937	2	US-08-847-429A-3	Sequence 3, Appli
C 41	30.8	2.1	937	5	US-09-065-474-1	Sequence 1, Appli
C 42	30.8	2.1	937	5	US-09-065-474-3	Sequence 3, Appli
C 43	30.8	2.1	1029	2	US-09-031-485-6	Sequence 6, Appli
C 44	30.8	2.1	1029	2	US-09-031-485-8	Sequence 8, Appli
C 45	30.8	2.1	1029	2	US-08-847-429A-6	Sequence 6, Appli

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22131-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14



```
;
; LENGTH: 480 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 162
; OTHER INFORMATION: /label= n
; OTHER INFORMATION: /note= "n-x-Unknown nucleotide"
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 356
; OTHER INFORMATION: /label= n
; OTHER INFORMATION: /note= "n-x-Unknown nucleotide"
;
US-08-282-581-4

Query Match          2.5%; Score 35.8; DB 1; Length 480;
Best Local Similarity 46.1%; Pred. No. 0.14;
Matches 118; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

Qy 1182 atattcatatatacaaaaacagagagagatccttaacaatacagagagcccttgagaatgga 1241
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 444 ATATATATATATAAAAAAGACAAAGAGGATCTCTCAGCAATTTCTCACCCAAAATTGA 385

Qy 1242 aacttggatcatgaatagtcctcttgaggatccttaacaatacagagagcccttgagaatgga 1301
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 384 TTACTTTATTAATAAATACTACTCTTNGCTGTTTGGTGACACCAAGCTAAAA 325

Qy 1302 ctagctttggccacacatcttttcatcaacagtttttaataatcgtgaggggagagaga 1361
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 TCCAATGTACCTGCTCTTTTATATATATATATATATATATATATATATATATATATA 265

Qy 1362 tggagataactaataatggacgtctattatagtttggaggtttttgtttttattttaa 1421
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 264 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 205

Qy 1422 tttagtaaaaaa 1437
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 204 AAATTTTAAAAACACA 189

RESULT 4
US-08-550-544-4/C
; Sequence 4, Application US/08550544
; Patent No. 5689056
; GENERAL INFORMATION:
; APPLICANT: Cramer, Carole L.
; TITLE OF INVENTION: HM2 PROMOTER EXPRESSION SYSTEM
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/550,544
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/100,816
; FILING DATE: 02-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
```

```
;
; REGISTRATION NUMBER: 18872
; REFERENCE/DOCKET NUMBER: 7956-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 162
; OTHER INFORMATION: /label= n
; OTHER INFORMATION: /note= "n-x-Unknown nucleotide"
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 356
; OTHER INFORMATION: /label= n
; OTHER INFORMATION: /note= "n-x-Unknown nucleotide"
;
US-08-550-544-4

Query Match          2.5%; Score 35.8; DB 1; Length 480;
Best Local Similarity 46.1%; Pred. No. 0.14;
Matches 118; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

Qy 1182 atattcatatatacaaaaacagagagagatccttaacaatacagagagcccttgagaatgga 1241
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 444 ATATATATATATAAAAAAGACAAAGAGGATCTCTCAGCAATTTCTCACCCAAAATTGA 385

Qy 1242 aacttggatcatgaatagtcctcttgaggatccttaacaatacagagagcccttgagaatgga 1301
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 384 TTACTTTATTAATAAATACTACTCTTNGCTGTTTGGTGACACCAAGCTAAAA 325

Qy 1302 ctagctttggccacacatcttttcatcaacagtttttaataatcgtgaggggagagaga 1361
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 TCCAATGTACCTGCTCTTTTATATATATATATATATATATATATATATATATATATA 265

Qy 1362 tggagataactaataatggacgtctattatagtttggaggtttttgtttttattttaa 1421
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 264 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 205

Qy 1422 tttagtaaaaaa 1437
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 204 AAATTTTAAAAACACA 189

RESULT 5
US-08-867-030B-5
; Sequence 5, Application US/08867030B
; Patent No. 5948900
; GENERAL INFORMATION:
; APPLICANT: Iother et al.
; TITLE OF INVENTION: Streptococcus pneumoniae
; TITLE OF INVENTION: Capsular Polysaccharide Genes and Flanking Regions
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benjamin Aaron Adler, Ph.D., J.D.
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch floppy
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
```

```

, APPLICATION NUMBER: US/08/867,030B
, FILING DATE: June 2, 1997
, CLASSIFICATION: 435
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: 08/243,546
, FILING DATE: May 16, 1994
, ATTORNEY/AGENT INFORMATION:
, NAME: Benjamin Aaron Adler, Ph.D., J.D.
, REGISTRATION NUMBER: 35,423
, REFERENCE/DOCKET NUMBER: D5923
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (713) 777-2321
, TELEFAX: (713) 777-6908
, INFORMATION FOR SEQ ID NO: 5:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 4951 bp
, TYPE: nucleic acid
, STRANDEDNESS: single-stranded
, TOPOLOGY: linear
, MOLECULE TYPE:
, DESCRIPTION: other nucleic acid
, HYPOTHETICAL: no
, ANTI-SENSE: no
, FRAGMENT TYPE:
, ORIGINAL SOURCE:
, IMMEDIATE SOURCE:
, POSITION IN GENOME:
US-08-867-030B-5

```

Query Match 2.4%; Score 34.8; DB 4; Length 4951;  
Best Local Similarity 46.3%; Pred. No. 1.2;  
Matches 114; Conservative 0; Mismatches 132; Indels 0

```

RESULT 6
PCT-US95-06119-5
: Sequence 5, Application PC/TUS9506119
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE CAPSULAR
: TITLE OF INVENTION: POLYSACCHARIDE GENES AND FLANKING REGIONS
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: TX
: COUNTRY: United States of America
: ZIP: 77210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06119
; FILING DATE: CONCURRENTLY HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/243,546
; FILING DATE: 16-MAY-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: AMCV018P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4951 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; PCT-US95-06119-5

```

Query Match	2.4k;	Score	34.8;	DB	6;	Length	4951;
Best Local Similarity	46.3k;	Pred.	No.	1.2;			
Matches	114;	Conservative	0;	Mismatches	132;	Indels	0;
Qy	1098	tggcgctgtgcaattaggagacattcatatacaaacaggaggtgattggttatactatttg	1157				
Db	3340	TTGACAGATGCTATTGATACATTGAATAAGACACAGAGTGTGTTTGGCGGTGAATTTCTG	3599				
Qy	1158	gccttttgcctgcgtgtgttggaaattcatatatacaaaaaacagaggaaggaacccctaac	1217				
Db	3600	GGCAAAAGCTTACGATGTGGTGATTAAGTTTAATTTTATGAAGAACATCAATTTGATTATGCT	3659				
Qy	1218	aatacagagagcccttggaaatgggaacttggatcatgaatatagcttcccttggaggatcaa	1277				
Db	3660	CTTCAACATCCTCAGATTAAAGACAGTTAAAAAATTTACGTTATTGGCACTTTGGTAACCAA	3719				
Qy	1278	gatgacacaccaagaaacccatagctagcttggcccaacaatccttttcatcaaacagtttta	1337				
Db	3720	TTGGAGAAGCTAGATGACTGTCTCAAGTGGACACCTATGAATTTGTATAGAAAGTTATC	3779				
Qy	1338	ataaat	1343				
Db	3780	AAAAAT	3785				

```

RESULT 7
US-08-370-975B-12
; Sequence 12, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
; APPLICANT: Maley, Gladys F.
; APPLICANT: Welner, Karen X.B.
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:

```

```

1  APPLICANT: Wellens, Thomas E.
2
3  TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
4  TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
5
6  NUMBER OF SEQUENCES: 45
7
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: Knobbe Martens Olson & Bear
10 STREET: 620 Newport Center Drive 16th Floor
11 CITY: Newport Beach
12 STATE: California
13 COUNTRY: US
14 ZIP: 92660
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: PatentIn Release #1.0, Version #1.25
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/08/487,826B
23 FILING DATE: 10-SEP-1993
24 CLASSIFICATION: 435
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Israelsen, Ned
27
28 REGISTRATION NUMBER: 29,655
29 REFERENCE/DOCKET NUMBER: NIH121.001CPI
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: (619) 235-8550
32 TELEFAX: (619) 235-0176
33 INFORMATION FOR SEQ ID NO: 13:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 19124 base pairs

```









Result No.	Score	Query			DB	ID	Description
		Match	Length	Percentage			
1	1437	100.0	1437	1	X02830	W09505387 Seq ID 5	
C 2	35.8	2.5	480	1	X081474	HMG2 promoter II.	
C 3	35.8	2.5	501	1	N50027	Sequence encoding	
C 4	35.6	2.5	501	1	N50029	Sequence encoding	
5	34.8	2.4	4951	1	T05848	Polycistronic cps	
6	34.6	2.4	4093	1	Q49264	ced-4. Long-distanc	
C 7	34.4	2.4	1755	1	T67326	Pasteurella haemol	
8	34	2.4	7778	1	V74359	Staphylococcus aur	
9	33.8	2.4	11764	1	X12985	Enterococcus faeca	
C 10	33.4	2.3	501	1	N50028	Sequence encoding	
C 11	33.4	2.3	110000	1	V12109_13	Continuation (14 o	
12	33.2	2.3	1078	1	X20337	Borrelia burgdorfer	
13	33	2.3	26764	1	T71696	Human deoxyxycidyl	
14	32.8	2.3	2079	1	X07338	Streptococcus pneu	
15	32.8	2.3	2339	1	T98581	DNA encoding a S.	
16	32.8	2.3	2511	1	X07337	Streptococcus pneu	
C 17	32.8	2.3	7913	1	V45588	Plasmid pUC-CONTR	
18	32.8	2.3	12835	1	V52274	Streptococcus pneu	
C 19	32.8	2.3	12845	1	V84274	Mouse endothelial	
20	32.8	2.3	19124	1	T72882	Plasmodium var-7 g	
C 21	32.6	2.3	501	1	N50024	Sequence encoding	
C 22	32.6	2.3	1579	1	X27359	Human secreted pro	
23	32.6	2.3	6146	1	X03045	Human IL-1ra BAC c	
C 24	32.6	2.3	6727	1	X02993	Human IL-1ra BAC c	
C 25	32.4	2.3	11225	1	N91695	Sequence of fowlpo	
C 26	32.4	2.3	110000	1	X20248_03	Continuation (4 of	
C 27	32.2	2.2	2941	1	Q52490	Helminth aminopept	
C 28	32.2	2.2	3358	1	Q52501	Helminth aminopept	
C 29	32.2	2.2	5189	1	V65701	DNA fragment of a	
30	32.2	2.2	9834	1	V74348	Staphylococcus aur	
C 31	32.2	2.2	1210000	1	V12109_14	Continuation (15 o	
32	32	2.2	729	1	N91692	Part of sequence o	
33	32	2.2	3974	1	T43360	Cotton FbLate2-82A	



```

RESULT 3
N50027/c
AC N50027 standard; DNA; 501 BP.
DE 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
DE IFNX 444.
KW Antiviral; cell growth regulator; Immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..501
/*tag= a
EP-163993-A.
PN 11-DEC-1985.
PD 17-MAY-1985; 105750.
PF (SEAR ) SEARLE G D & CO.
PR 17-MAY-1984; GB-012564.
PA Bell LD, Boseley PG, Porter AG;
PI WPI: 85-311944/50.
DR P-PSDB; P50026.
DR P-PSDB; P50026.
PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2e, page 36; 7lpp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 Bp; 112 A; 31 C; 67 G; 80 T;

Query Match 2.5%; Score 35.8; DB 1; Length 501;
Best Local Similarity 28.8%; Pred. No. 0.86;
Matches 61; Conservative 50; Mismatches 101; Indels 0; Gaps 0;

Qy 796 tcttgtcacaacaggctctattttgtccacacacagcactgttggtcagaataactgcg 855
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Dy 239 TTCCANCCNGTNNNNNNNNWTCYTGCKRAADATNGCRAADATRTTYTGVARCATYTCR 180
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Qy 856 tgaagagaaagatcgattgaagtagtagcaatgacgagcctcctcactcaatgcagtggtcactca 915
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Dy 179 TADATNGTVARNGCGRCTCYTCYTTATGAAATGTYTGNCCTGYTTDATYTCYTCNGGD 120
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Qy 916 gtcaacccgagattactgtctgagagagaaatgcctcctcactcaatgcagtggtcactg 975
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Dy 119 ATNCCRAATTTCATHCRTCTYTVARRCARTAYTCVARHCKNCCTTVARYTGCCAVARV 60
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Qy 976 gactttggcagcctatgtgtttatgcactg 1007
:||||||| | : : : : : | : : : : : | : : : : : | : : : : :
Dy 59 ARYTTYTGRCAYTGAAATNNNNNNWHCKYTG 28
||||| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

RESULT 4
N50029/c
AC N50029 standard; DNA; 501 BP.
DE 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
DE IFNX 446.
KW Antiviral; cell growth regulator; Immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..501
/*tag= a
EP-163993-A.
PN 11-DEC-1985.
PD 17-MAY-1985; 105750.
PF (SEAR ) SEARLE G D & CO.
PR 17-MAY-1984; GB-012564.
PA Bell LD, Boseley PG, Porter AG;
PI WPI: 85-311944/50.
DR P-PSDB; P50026.
DR P-PSDB; P50026.
PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2e, page 36; 7lpp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 Bp; 112 A; 31 C; 67 G; 80 T;

Query Match 2.5%; Score 35.8; DB 1; Length 501;
Best Local Similarity 28.8%; Pred. No. 0.86;
Matches 61; Conservative 50; Mismatches 101; Indels 0; Gaps 0;

Qy 796 tcttgtcacaacaggctctattttgtccacacacagcactgttggtcagaataactgcg 855
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Dy 239 TTCCANCCNGTNNNNNNNNWTCYTGCKRAADATNGCRAADATRTTYTGVARCATYTCR 180
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Qy 856 tgaagagaaagatcgattgaagtagtagcaatgacgagcctcctcactcaatgcagtggtcactca 915
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Dy 179 TADATNGTVARNGCGRCTCYTCYTTATGAAATGTYTGNCCTGYTTDATYTCYTCNGGD 120
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Qy 916 gtcaacccgagattactgtctgagagagaaatgcctcctcactcaatgcagtggtcactg 975
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Dy 119 ATNCCRAATTTCATHCRTCTYTVARRCARTAYTCVARHCKNCCTTVARYTGCCAVARV 60
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Qy 976 gactttggcagcctatgtgtttatgcactg 1007
:||||||| | : : : : : | : : : : : | : : : : : | : : : : :
Dy 59 ARYTTYTGRCAYTGAAATNNNNNNWHCKYTG 28
||||| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

RESULT 5
T05848
ID T05848 standard; DNA; 4951 BP.
AC T05848;
DT 03-APR-1996 (first entry)
DE Polyclonistic cps gene locus operon of S.pneumoniae.
KW Capsular polysaccharide; cps; peptide; flanking region; detection;
KW serotype; diagnosis; prevention; Streptococcus pneumoniae; ss.
OS Streptococcus pneumoniae.
FH Key Location/Qualifiers
FT cds 1..1277
/*tag= a
/*product= Capsular polysaccharide CpsD.
FT cds 1277..2543
/*tag= b
/*product= Capsular polysaccharide CpsS.
FT cds 2707..3761
/*tag= c
/*product= Capsular polysaccharide CpsU.
FT cds 3758..4951
/*tag= d
/*product= Capsular polysaccharide CpsM.
W09531548-A1.
PN 23-NOV-1995.
PD 16-MAY-1995; U06119.
PF 16-MAY-1994; US-243546.
PR (UABR-) UAB RES FOUND.
PA Dillard JP, Yother J;
PI 96-010934/01.
DR P-PSDB; R83038, R83039, R83040, R83041.
DR New Streptococcus pneumoniae capsular polysaccharide genes - used
PT for detection, serotyping and for diagnosis and prevention of S.
PT pneumoniae infection
PT Disclosure; Page 156-163; 226pp; English.

```

```

PR 17-MAY-1984; GB-012564.
PA (SEAR ) SEARLE G D & CO.
PI Bell LD, Boseley PG, Porter AG;
DR WPI: 85-311944/50.
DR P-PSDB; P50028.
DR P-PSDB; P50028.
PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2g, page 38; 7lpp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 Bp; 112 A; 31 C; 69 G; 79 T;

Query Match 2.5%; Score 35.6; DB 1; Length 501;
Best Local Similarity 28.8%; Pred. No. 0.98;
Matches 61; Conservative 49; Mismatches 102; Indels 0; Gaps 0;

Qy 796 tcttgtcacaacaggctctattttgtccacacacagcactgttggtcagaataactgcg 855
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Dy 239 TTCCANCCNGTNNNNNNNNWTCYTGCKRAADATNGCRAADATRTTYTGVARCATYTCR 180
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Qy 856 tgaagagaaagatcgattgaagtagtagcaatgacgagcctcctcactcaatgcagtggtcactca 915
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Dy 179 TADATNGTVARNGCGRCTCYTCYTTATGAAATGTYTGNCCTGYTTDATYTCYTCNGGD 120
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Qy 916 gtcaacccgagattactgtctgagagagaaatgcctcctcactcaatgcagtggtcactg 975
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Dy 119 ATNCCRAATTTCATHCRTCTYTVARRCARTAYTCVARHCKNCCTTVARYTGCCAVARV 60
||| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Qy 976 gactttggcagcctatgtgtttatgcactg 1007
:||||||| | : : : : : | : : : : : | : : : : : | : : : : :
Dy 59 ARYTTYTGRCAYTGAAATNNNNNNWHCKYTG 28
||||| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

RESULT 5
T05848
ID T05848 standard; DNA; 4951 BP.
AC T05848;
DT 03-APR-1996 (first entry)
DE Polyclonistic cps gene locus operon of S.pneumoniae.
KW Capsular polysaccharide; cps; peptide; flanking region; detection;
KW serotype; diagnosis; prevention; Streptococcus pneumoniae; ss.
OS Streptococcus pneumoniae.
FH Key Location/Qualifiers
FT cds 1..1277
/*tag= a
/*product= Capsular polysaccharide CpsD.
FT cds 1277..2543
/*tag= b
/*product= Capsular polysaccharide CpsS.
FT cds 2707..3761
/*tag= c
/*product= Capsular polysaccharide CpsU.
FT cds 3758..4951
/*tag= d
/*product= Capsular polysaccharide CpsM.
W09531548-A1.
PN 23-NOV-1995.
PD 16-MAY-1995; U06119.
PF 16-MAY-1994; US-243546.
PR (UABR-) UAB RES FOUND.
PA Dillard JP, Yother J;
PI 96-010934/01.
DR P-PSDB; R83038, R83039, R83040, R83041.
DR New Streptococcus pneumoniae capsular polysaccharide genes - used
PT for detection, serotyping and for diagnosis and prevention of S.
PT pneumoniae infection
PT Disclosure; Page 156-163; 226pp; English.

```





```
Query Match          2.4%; Score 33.8; DB 1; Length 11764;
Best Local Similarity 54.4%; Pred. No. 15;
Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 171 actctcgaagaagaataataataataatggcgatagatctctttagatagatgtaggt 230
    || || || || || || || || || || || || || || || || || || || ||
Db 73 AAGATTAAATGAATAAATAACAAGCAAAATATAAGATTTTTTTGTTTAACTCTTTTGT 132

Qy 231 gtactctgcagggtgcttaataatctccggtggtgggaaataataataataataata 290
    || || || || || || || || || || || || || || || || || || || ||
Db 133 TTTCAATAGCAAGTGTCTAGTGTGTTTTTCAGAAATTGAAAAATGAATACGACAAAAATTTA 192

Qy 291 gcgat 295
    || ||
Db 193 TCTAT 197

RESULT 10
N50028/c
ID N50028 standard; DNA; 501 BP.
AC N50028.
DE 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
KW Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..501
FT /tag= a
PN EP-163993-A.
PD 11-DEC-1985.
PF 17-MAY-1985; 105750.
PR 17-MAY-1984; GB-012564.
PA (SEAR ) SEARLE G D & CO.
PI Bell LD, Boseley PG, Porter AG;
DI WPI; 85-311944/50.
DR P-PSDB; P50027.
PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2f, page 37; 71pp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 BP; 114 A; 31 C; 68 G; 79 T;

Query Match          2.3%; Score 33.4; DB 1; Length 501;
Best Local Similarity 28.3%; Pred. No. 4.1;
Matches 60; Conservative 49; Mismatches 103; Indels 0; Gaps 0;

Qy 796 tctctgtcaaacagcgtctatttggttcaacactcagcactgttggtcaggaaactgog 855
    || || || || || || || || || || || || || || || || || || || ||
Db 239 TTTCCANCCNGTNNWNNWNTCTYTHCKRAADATNGCRAADATRTTYTGVARCATYTCR 180

Qy 856 tgaagaggaaagatcgatgaagtagtagcaatgacgtggtattgttgatgctcatca 915
    || || || || || || || || || || || || || || || || || || || ||
Db 179 TADATNGIVARNGCNGRCYCTCYTTTGTGRAAYTGTCNCCYGTGTTDATYTCYTCNGD 120

Qy 916 gtgcaccgagattactgtctgagaggagaatgccctctcatcaatgcagtggtctactg 975
    || || || || || || || || || || || || || || || || || || || ||
Db 119 APTTCRAARTTCAHKCTCYTTVARRCARTAYTCVARHCKNCCRTTVARYTGCCAVARV 60

Qy 976 gacttttgagcactatgttggttgactg 1007
    || || || || || || || || || || || || || || || || || || || ||
Db 59 ARATTTGRCAYTGTAARTNNWNNWHCTYTG 28
```

```
RESULT 11
V21209_13/c
Continuation (14 of 17) of V21209 from base 1300001 (Methanococcus jannaschlii circula
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
WP Fragment Name Begin End
WP V21209_00 1 110000
WP V21209_01 100001 210000
WP V21209_02 200001 310000
WP V21209_03 300001 410000
WP V21209_04 400001 510000
WP V21209_05 500001 610000
WP V21209_06 600001 710000
WP V21209_07 700001 810000
WP V21209_08 800001 910000
WP V21209_09 900001 1010000
WP V21209_10 1000001 1110000
WP V21209_11 1100001 1210000
WP V21209_12 1200001 1310000
WP V21209_13 1300001 1410000
WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976

Query Match          2.3%; Score 33.4; DB 1; Length 110000;
Best Local Similarity 52.5%; Pred. No. 55;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 151 gtactgtactactgactcaactctcaagagaagaataataataataatggcgatag 210
    || || || || || || || || || || || || || || || || || || || ||
Db 33193 GTACAGCTATCAACGATGAATTAATAAGACAATTGTTAAATTAGCTAAACAGCTAAAC 33134

Qy 211 atctttgtagatcatgtaggtgtagcctgcaggtggttaataatctccggtgtgggaaaa 270
    || || || || || || || || || || || || || || || || || || || ||
Db 33133 AAATTTGTTAAATTCTTCAACAAAGACAGAGAGAGTAAATTTTATCATCCCTGCTTATACA 33074

Qy 271 taataataataataataat 289
    || || || || || || || || || || || || || || || || || || || ||
Db 33073 GAGGTTGATGAAGTAATAT 33055

RESULT 12
X20337
ID X20337 standard; DNA; 1078 BP.
AC X20337.
DT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #90.
DE Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN WO9858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
DR WPI; 98-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 1072; 1128pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
```



PI recombinant deaminase  
PS Claim 3: Column 55-78: 58pp: English

CC treatment to enhance (agonist) or blunt (antagonist) SecA activity, thereby treating the cells with a SecA agonist or antagonist.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: August 1, 2000, 13:21:41 ; Search time 9137.45 seconds  
(without alignments)  
280.670 Million cell updates/sec

Title: US-09-147-955-9  
Perfect score: 1437  
Sequence: 1 ttcaaaactcataacgtgat.....tttatattgagtaaaaaaaa 1437

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues  
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: em\_fun:\*  
17: em\_humi:\*  
18: em\_hum2:\*  
19: em\_in:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_sy:\*  
29: em\_un:\*  
30: em\_vi:\*  
31: gb\_htg1:\*  
32: gb\_htg2:\*  
33: gb\_in1:\*  
34: gb\_in2:\*  
35: em\_ba1:\*  
36: em\_ba2:\*  
37: em\_hum3:\*  
38: em\_hum4:\*  
39: gb\_pr4:\*  
40: gb\_htg3:\*  
41: gb\_htg4:\*  
42: gb\_htg5:\*  
43: gb\_htg6:\*

44: gb\_htg7:\*  
45: em\_htg1:\*  
46: em\_htg2:\*  
47: em\_htg3:\*  
48: em\_hums:\*  
49: gb\_pl3:\*  
50: gb\_pr5:\*  
51: gb\_htg8:\*  
52: gb\_htg9:\*  
53: gb\_htg10:\*  
54: gb\_htg11:\*  
55: gb\_htg12:\*  
56: gb\_htg13:\*  
57: gb\_htg14:\*  
58: gb\_in3:\*  
59: gb\_htg15:\*  
60: gb\_htg16:\*  
61: gb\_htg17:\*  
62: em\_htg4:\*  
63: em\_htg5:\*  
64: em\_htg6:\*  
65: em\_htg7:\*  
66: em\_hum6:\*  
67: gb\_htg18:\*  
68: gb\_htg19:\*  
69: gb\_htg20:\*  
70: gb\_htg21:\*  
71: gb\_htg22:\*  
72: gb\_htg23:\*  
73: gb\_htg24:\*  
74: gb\_htg25:\*  
75: gb\_htg26:\*  
76: gb\_htg27:\*  
77: gb\_htg28:\*  
78: gb\_htg29:\*  
79: gb\_htg30:\*  
80: gb\_htg31:\*  
81: gb\_vil:\*  
82: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	92.8	6.5	40005	7	SPCC320	AL022245 S.pombe c
C 2	45	3.1	7218	5	I66494	I66494 Sequence 14
C 3	43.8	3.0	65870	12	AC005807	AC005807 Mus muscu
C 4	43.6	3.0	7218	5	I66494	I66494 Sequence 14
C 5	43	3.0	96958	39	AC006350	AC006350 Homo sapi
C 6	42.4	3.0	98630	49	AC006535	AC006535 Genomic s
C 7	42.2	2.9	1857	12	MUSMTL1	M11742 Mouse MHC C
C 8	42.2	2.9	5187	12	MUSMHLA	M75875 Mouse MHC C
C 9	41.6	2.9	148241	41	AC009857	AC009857 Homo sapi
C 10	41.6	2.9	167429	70	AC010952	AC010952 Homo sapi
C 11	41.6	2.9	196085	32	CNS01RGW	AL160231 Homo sapi
C 12	41.2	2.9	200337	61	AC021186	AC021186 Homo sapi
C 13	41	2.9	188461	42	AC010949	AC010949 Homo sapi
C 14	41	2.9	189910	40	CNS01RG0	AL157888 Homo sapi
C 15	40.8	2.8	22696	42	AC012895	AC012895 Drosophil
C 16	40.8	2.8	156492	54	AC008460	AC008460 Homo sapi
C 17	40.8	2.8	218596	60	AC021781	AC021781 Homo sapi
C 18	40.8	2.8	327311	34	AE003500	AE003500 Drosophil
C 19	40.6	2.8	102263	32	AL139152	AL139152 Homo sapi
C 20	40.6	2.8	167101	60	AC015595	AC015595 Homo sapi
C 21	40.4	2.8	84412	10	HS973M2	AL033533 Human DNA
C 22	40.4	2.8	101154	8	ATF1P2	AL132955 Arabidops
C 23	40.2	2.8	158546	54	AC022876	AC022876 Homo sapi
C 24	40.2	2.8	181671	31	AP001024	AP001024 Homo sapi

```

25 40.2 2.8 200038 43 AC018423
26 40 2.8 58834 51 AC022723
27 40 2.8 165048 54 AC012101
28 40 2.8 176008 40 AL353113
29 39.8 2.8 456 9 HSN0V1
30 39.8 2.8 59654 57 AC025727
31 39.8 2.8 110000 31 PFNAL13P2_2
32 39.6 2.8 85992 7 AB009052
33 39.4 2.7 76175 49 AC009273
34 39.4 2.7 97189 10 HS479J7
35 39.2 2.7 114534 31 AP000633
36 39.2 2.7 152981 31 AP001193
37 39.2 2.7 159164 31 AP001140
38 39.2 2.7 172078 40 AL160276
39 39.2 2.7 183461 39 AC006121
40 39 2.7 159553 44 AC009990
41 39 2.7 209826 74 AC021130
42 38.8 2.7 8558 33 CEG15074
43 38.8 2.7 29350 33 CEG14B1
44 38.8 2.7 68794 72 AC034112
45 38.8 2.7 130514 60 AC016441

```

## ALIGNMENTS

```

RESULT 1
SPCC320/C 1
LOCUS SPCC320 40005 bp DNA PLN 06-APR-1999
DEFINITION S.pombe chromosome III cosmid c320.
ACCESSION AL022245
VERSION AL02245.2 GI:4584240
KEYWORDS beta-synthase; bio2, biotin synthase; EF-hand calcium-binding
domain; ferredoxin; protein kinase; putative transporter; ribosome
subunit biogenesis; rna binding protein; serine/threonine
dehydratase; signal recognition particle protein subunit;
transcriptional regulator; trna- Glu.
flission yeast.
SOURCE Schizosaccharomyces pombe
ORGANISM Eukaryota; Fungi; Ascomycota; Archiascomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
REFERENCE 1 (bases 1 to 40005)
AUTHORS Wood,V., Rajandream,M.A., Barrell,B.G., Wedler,H., Wambutt,R. and
Wedler,E.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1998) European Schizosaccharomyces genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
and AGOWA GmbH, Glienicke Weg 185, D-12489 Berlin, Germany
COMMENT On Apr 14, 1999 this sequence version replaced gi:2995362.
Notes:
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites. CDS are numbered using the following
system eg SPCC320.10c. SP (S. pombe), B (chromosome 2), c25H2
(cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Details of yeast
sequencing at the Sanger Centre are available on the World Wide
Web.
(URL, http://www.sanger.ac.uk/Projects/S.pombe/)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was

```

sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project. Cosmid c320 is overlapped by cosmid c330 at the 5' end and c1235 at the 3' end.

```

FEATURES             Location/Qualifiers
     source            1..40005
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="III"
                     /clone="cosmid c320"
                     /map="TIII"
     misc_feature      1..7687
                     /note="nominal overlap with cosmid SPCC330 S. pombe
                     chromosome 3"
     gene              complement(709..1680)
                     /gene="spcc320.14"
     CDS               complement(709..1680)
                     /gene="spcc320.14"
                     /note="SPCC320.14, len:322, SIMILARITY: Saccharomyces
                     cerevisiae, YKL218C, YKV8_YEAST, P36007, hypot helical
                     protein with strong similarity to E.coli and H.influenzae
                     threonine dehydratases, (326aa), fasta scores, opt:14 64,
                     E():0, (66.4% identity in 318 aa overlap)"
                     /codon_start=1
                     /product="putative serine/threonine dehydratase"
                     /protein_id="CAA18316.1"
                     /db_xref="GI:2995376"
                     /db_xref="SPTREMBL:O59791"
                     /translation="MSDNLVLPYDDVASERIKKFKPNTKPVLTSTSTVNFVAFV
                     FKCEQKMGAFKRGALNALSQLNEAQRKAGVLTFSSGNHQAIALSAILGIPAKI
                     IMPLDAPKFAVATKGGQVIMDRDKDRKMAKREISREGLTIPTDHPHVLG
                     QGTAAELFEVGLDGLFCVLGGGLLSGSAALAAHFAPNCEYGVGPEAGNDGQS
                     FRGSIVHDITPKTIADGAQTHLGNVTFSIKEKVDITLTVSDEELIDCLKPYAARM
                     KIVVEPTGCLSFAAARAKKELKRNKIGIIGSGNVDIERYAHFLSO"
                     complement(1123..1518)
                     /gene="spcc320.14"
                     /note="Pfam match to entry S_T_dehydratase ,
                     beta-synthase"
                     2454..2525
                     /gene="trna Glu"
                     /product="trna Glu"
                     2454..2525
                     /gene="trna Glu"
                     3062..4216
                     /gene="spcc330.16"
                     3062..4216
                     /gene="spcc330.16"
                     /note="SPCC330.16c, len:384, SIMILARITY: Xenopus laevis,
                     Q91820, P46EG265, The Xenopus protein kinase p82
                     associates with the centrosome in a cell cycle-dependent
                     manner, binds to the spindle microtubules and is involved
                     in bipolar mitotic spindle assembly., fasta scores,
                     opt:1185, E():0, (48.2% identity in 380 aa overlap), also
                     similar eg. to P97477, serine/threonine protein kinase,
                     (395aa), fasta scores, opt:1157, E():0, (54.9% identity in
                     317 aa overlap)"
                     /codon_start=1
                     /product="ser/thr protein kinase"
                     /protein_id="CAA18315.1"
                     /db_xref="GI:2995375"
                     /db_xref="SPTREMBL:O59790"
                     /translation="MVLQVNNNAPTSGRTCKPTSAFFYEIMSDSKLADSLNCLSV
                     STPTTANQRQQLRLAVSNQVNNVSLANGKNTSNKSNFNSLRKIEEPIAGV
                     PSSAGPQWREHFHICMEIGKLGKFGVRVLAKEKRTGFIIVALKTLHKSELVQSKTE
                     KOVRETEQSNLRNKLRLYGHFDEKRIYLLLEFAGRGELYOHLRAKRFSEVA
                     SKYIFOMANALSVLHKHNVHTRDIKPENILLIDGTEKLSDFGSHVAPNRRTLCG
                     TLYLDPENVEGHEKVDLSGLVITETFLVAGPFFEDMSGSHSATYTKRIAKVDLKI
                     PSFVPPDARDLISRLQHNPEKRMSLEQVNRHRPWIVKYKDSWTRKSSSS"
                     3413..4168
                     /gene="spcc330.16"
     misc_feature      3413..4168

```

```

/misc_feature
/Note="Pfam match to entry pkinase PF00069, Eukaryotic
prot ein kinase domain"
3431..3502
/Genes="spcc330.16"
/Note="PS00107 Protein kinases ATP-binding region
signature"
3770..3808
/Genes="spcc330.16"
/Note="PS00108 Serine/Threonine protein kinases
active-site signature"
complement(4707..5330)
/Genes="spcc320.12"
complement(join(4707..4714,4781..5330))
/Genes="spcc320.12"
/Note="SPCC320.12, len:184, SIMILARITY:Saccharomyces
cerevisiae, YN8020C, YN8A_YEAST,hypothetical protein,
(270aa), fast a scores, Opt:592, E():0, (47.5% identity in
183 aa overlap)"
/codon_start=1
/product="hypothetical 21.5 kd protein"
/protein_id="CAA18314.1"
/db_xref="GI:2995374"
/db_xref="SPTREMBL:O59789"
/translation="MSEDPQSMKSKKNCERVKRALMSQSPVILFKTALDRLCNIE
AKDISCPDAOSTGGYIPGKIGVILCENRLYTKKNAENTIAHEIMHMFDDHREVDWN
NLRHQACSEIRASSGSGRCWTKELRFGNIKTRKHQECVKRRATISVQGNPKSK
EQEALIVEVFNCSNDRPRPEKII"
complement(4715..4738)
/Genes="spcc320.12"
/Note="PS00142 Neutral zinc metalloproteinases,
zinc-binding region signature"
5867..6724
/Genes="SPCC320.11c"
Join(5867..5922,6076..6162,6325..6724)
/Genes="SPCC320.11c"
/Note="SPCC320.11c, len:179, SIMILARITY:Saccharomyces
cerevisiae, YP211W, Q08962, required for efficient 60S
ribosome subunit bio genesis, (181aa), fasta scores,
Opt:892, E():0, (70.6% identity in 180 aa overlap)"
/codon_start=1
/product="putative ribosome biogenesis protein"
/protein_id="CAA18313.1"
/db_xref="GI:2995373"
/db_xref="SPTREMBL:O59788"
/translation="MRLPHEETKTFEKLQAYIGKNITHLIDRPDPHCFRLQDRV
YVVSERAMKATSVARQNLMSLIGCFKFTKRLTALDYIAQYARYKIWKSN
GMPFLYGNHVLHGVRITDDTPQHQGVYVSMNDPLGFGVTARSTLELRLEPTA
IVAFHQADYGEVLRDEDTLF"
5923..5928
/Genes="SPCC320.11c"
/Note="gtaagt, splice donor sequence"
6059..6075
/Genes="SPCC320.11c"
/Note="ctgactagtaaaactag, splice branch and acceptor"
6163..6168
/Genes="SPCC320.11c"
/Note="SPCC320.11c"
6310..6324
/Genes="SPCC320.11c"
/Note="ttaacatacttag, splice branch and acceptor"
complement(7697..9615)
/Genes="SPCC320.10"
complement(join(7697..8738,8785..9104,9156..9218,
9355..9615))
/Genes="SPCC320.10"
/Note="SPCC320.10, len:560, SIMILARITY:Homo sapiens,

```

```

SR72_HUMAN, signal recognition particle 72 kd protein,
(670 aa), fasta scores: Opt: 354, E():1.4e-14, (24.6%
identity in 574 aa)
/codon_start=1
/label="SPCC320.10"
/product="putative signal recognition particle protein"
/protein_id="CAA18312.1"
/db_xref="GI:2995372"
/db_xref="SWISS-PROT:O59787"

Query Match      6.5%; Score 92.8; DB 7; Length 40005;
Best Local Similarity 47.4%; Pred. No. 5.4e-13;
Matches 314; Conservative 0; Mismatches 342; Indels 6; Gaps 1;

QY 537 tatgtctctcctgggggttcgtggcagctccacggcaattatcttggtaataagaattcgag 596
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13681 TATATCATCTTTGCCCTTTTGGAGTTGAAGGAAATATTTTCGTTGACTAGCCTACCAA 13622

QY 597 ttgacatcgattacgagtgtagcactactgattgttgagcaatcgtgtggtccatcatc 656
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13621 TACACAACATGCTTAGTGAAGTTTACTCGATTCTTGGGCTACTGTCGAGTTGTAAAT 13562

QY 657 ttacatggatgttccttag-----gcacaaatactctgtatataccagttgttcggtgct 710
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13561 CTTTCCCTTTATCTTTTAAAGGTTTCGCTATCCTCTGGTCTCAAAATCTTGGTGTCTGGCT 13502

QY 711 gctatttgttagaggcctcctcctcgtcgtcttcttccgactcaggggtcactgctgct 770
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13501 TGCATTGGTGTGTTGGTGGTGGTGTAGTGATGATTTCTCGTGGGTGATTAATCT 13442

QY 771 gggtcgaatcctctttgggtgatttcttctgataacagagctctatttggttcacactc 830
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13441 GCTGTCAACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 13382

QY 831 agcactgttgtcaggaaatactcgtggaaggaagaaatcgtattgaagtagtagcaatg 890
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13381 TCCAATACTTTGGAGGAATACTTTGCTTCCAAATTCGCGCTTTACGTAGTCTATCGGTCAA 13322

QY 891 atcgggtatttgtgtatgctcactcagtgcaacgagattactgctggagagaagaatgcc 950
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13321 TTAAGTCTCTACGGTCTTATATACAGCATATTTCAAACCTTTATTTTGACCGTCACCAT 13262

QY 951 ctctcatcaatgcaagtgttactggacttttggcagcctatgtgtttatgcactgcc 1010
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13261 TTGTATACCTCGCATTCGACTAGTGAATGGGTGGTACTTAGCTGGTGGTGGTGGTGGTGG 13202

QY 1011 agcttcctcttgcacactcaacccctttcttctcaagatgagtggtgcatttttc 1070
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13201 ATGTTCTTGCTTTATAGTTTGGCTCCCTATTCTCTTTAGAAATGCTCTCTCGACATTTTAT 13142

QY 1071 aatcttccatgcttaccatctgatatgtgggctgttgcattgttgcattatcatacaac 1130
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13141 AATATCTCTTTATTTGACTTCCGATTTTGGAGCTTGGTATTATTTGTTATTCATGTGTTTGA 13082

QY 1131 caggaggttcaatgggttatactatttggccttttctcgttctgttgggaattcata 1190
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13081 TATCATGTGACTGGCTGTACCCCATTCCTTCGTCTTAATATATATATATATATATATATAT 13022

QY 1191 ta 1192
|||
Db 13021 TA 13020

RESULT 2
LOCUS 166494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.

```



```
Qy 202 gggcgatagatcttgtagatagtgtagctgagcagtggtgtaataatcccggt 261
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1328 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1269
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 262 gtgggaaataaataaataatagcagcagcagcagcagcagcagcagcagcagc 321
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1268 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1209
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 322 gggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 371
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1208 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1159
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
AC006350
LOCUS AC006350 96958 bp DNA PRI 21-DEC-1999
DEFINITION Homo sapiens PAC clone RP5-998H4 from 7q21.1-q21.2, complete
          sequence.
ACCESSION AC006350
VERSION AC006350.2 GI:4508140
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 96958)
AUTHORS Sulston J.E. and Waterston R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 96958)
AUTHORS Madsen C. and Spalding L.
TITLE The sequence of Homo sapiens PAC clone RP5-998H4
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 96958)
AUTHORS Waterston R.H.
TITLE Direct Submision
JOURNAL Submitted (11-JAN-1999) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 4 (bases 1 to 96958)
AUTHORS Waterston R.H.
TITLE Direct Submision
JOURNAL Submitted (24-MAR-1999) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 5 (bases 1 to 96958)
AUTHORS Waterston R.
TITLE Direct Submision
JOURNAL Submitted (28-APR-1999) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          (bases 1 to 96958)
REFERENCE 6 (bases 1 to 96958)
AUTHORS Waterston R.
TITLE Direct Submision
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          On Mar 24, 1999 this sequence version replaced gi:4139372.
COMMENT -----
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu/gsc
          Contact: sapiens@watson.wustl.edu
          -----
          Summary Statistics
          -----
          Center project name: H_DJ0998H04
          -----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is CTB-22J17, 200 bp overlap. Actual end of this clone is at base position 1 of RP5-998H4; actual end is at 27949 of CTB-22J17.

FEATURES

Location/Qualifiers	Source
1..96958	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7q21.1-q21.2"
	/clone="RP5-998H4"
	/clone_lib="RPCI-5"
799..881	/rpt_family="L1"
882..1242	/rpt_family="MaLR"
1243..2444	/rpt_family="MaLR"
2445..2753	/rpt_family="Alu"
2754..3185	/rpt_family="MaLR"
3186..3547	/rpt_family="MaLR"
3548..3652	/rpt_family="L1"
4058..4356	/rpt_family="Alu"
4708..4991	/rpt_family="Alu"
4992..5017	/rpt_family="(TAA)n"
5741..5810	/rpt_family="L2"
5842..6141	/rpt_family="Alu"
6313..8075	/rpt_family="L1"
8228..8268	/rpt_family="(CA)n"
8772..8865	/rpt_family="MER21_g"
8900..9034	/rpt_family="MIR"
9742..9866	/rpt_family="MIR"
10829..10874	

```

/rpt_family="(TGA)n"
11577..11858
/rpt_family="Alu"
11859..11905
/rpt_family="(TAA)n"
11997..12295
/rpt_family="Alu"
12857..12938
/rpt_family="L2"
13282..13487
/rpt_family="MIR"
14527..14959
/rpt_family="MER2_type"
15116..15187
/rpt_family="MIR"
15812..15932
/rpt_family="(TA)n"
15853..15972
/rpt_family="(CATA)n"
15994..16110
/rpt_family="(CATA)n"
16108..16167
/rpt_family="(CATA)n"
16277..16414
/rpt_family="MER1_type"
16429..17076
/rpt_family="L1"
17128..17251
/rpt_family="L1"
17667..17727
/rpt_family="(CTATT)n"
18293..18556
/rpt_family="L2"
19543..19579
/rpt_family="AT-rich"
19582..19611
/rpt_family="(TAAA)n"
19612..19668
/rpt_family="Alu"
20150..20176
/rpt_family="AT-rich"
20685..20746
/rpt_family="(CA)n"
20928..21112
/rpt_family="L2"
21236..21305
/rpt_family="MER1_type"
21345..21386
/rpt_family="AT-rich"
21396..21515
/rpt_family="L2"
22491..22521
/rpt_family="AT-rich"
22882..22935
/rpt_family="Other"
23339..23724
/rpt_family="L2"
23821..23938
/rpt_family="Other"
23983..24209
/rpt_family="Retroviral"
24210..24286
/rpt_family="MER4-group"
24387..25268
/rpt_family="Retroviral"
25276..26438
/rpt_family="L1"
26435..26703
/rpt_family="L1"

```

Query Match 3.08; Score 43; DB 39; Length 96958;  
 Best Local Similarity 50.79; Pred. No. 1.8;  
 Matches 103; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```

QY 1234 agaatggaacttgatcatgaatagctctcttgagatcaaatgacacacacagaa 1293
DB 69232 ACAGTAGACAATAGGGTCATCAAAATGAAACCGCTAATGCGCAAGGACACAA 69291
QY 1294 aaccatagctagctttgcccacaatcttttcatcaacagtttttaataatctgtagggg 1353
DB 69292 TAAGCTTTCTTTTCAAAAAATGGTTTCATCAGCATTTCTAATGACATCTCAAGGCT 69351
QY 1354 gagagatcgagatactactaataatggacgtctattatattagttggaggtttttttta 1413
DB 69352 TATAGTAAAGGTAGACCAAAATAGCTAGGTGGAGTGTGTTAAATGCTTTTATTGT 69411
QY 1414 tttattatttgagtaaaaaaaa 1436
DB 69412 TACAATTAATTTCAAGTTAAAA 69434

RESULT 6
AC006535 98630 bp DNA PLN 15-FEB-2000
LOCUS Genomic sequence for Arabidopsis thaliana BAC T24P13 from
DEFINITION Chromosome I, complete sequence.
ACCESSION AC006535
VERSION AC006535.7 GI:6978061
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 98630)
AUTHORS Johnson-Hopson,C., Dunn,P., Brooks,S., Buehler,E., Chao,Q.,
Khan,S., Kim,C., Shinn,P., Altafi,H., Bei,Q., Chin,C., Chiou,J.,
Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B.,
Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J.,
Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G.,
Federspiel,N.A., Theologis,A. and Ecker,J.R.
Genomic sequence for Arabidopsis thaliana BAC T24P13 from
Chromosome I
Unpublished
REFERENCE 2 (bases 1 to 98630)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 3 (bases 1 to 98630)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 4 (bases 1 to 98630)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
COMMENT On Feb 15, 2000 this sequence version replaced gi:4803878.
FEATURES
source
1. 98630
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="T24P13"

BASE COUNT 32079 a 17075 c 16743 g 32733 t
ORIGIN

```

Query Match 3.0%; Score 42.4; DB 49; Length 98630;



matches vv, conservative v, misalliances vv, tunnels vv, gaps

```

exon1
      2011..3060
      /gene="MHC class I T3-d"

```



Lehoczyk, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Melchior, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Rhoman, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 30, 2000 this sequence version replaced gi:6249728.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L2836  
Center clone name: 4\_N16  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 41 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1114: contig of 1114 bp in length  
\* 1115 1214: gap of 100 bp  
\* 1215 2323: contig of 1109 bp in length  
\* 2324 2423: gap of 100 bp  
\* 2424 3499: contig of 1076 bp in length  
\* 3500 3599: gap of 100 bp  
\* 3600 4651: contig of 1092 bp in length  
\* 4692 4791: gap of 100 bp  
\* 4792 6613: contig of 1822 bp in length  
\* 6614 6713: gap of 100 bp  
\* 6714 8003: contig of 1290 bp in length  
\* 8004 8103: gap of 100 bp  
\* 8104 9521: contig of 1418 bp in length  
\* 9522 9621: gap of 100 bp  
\* 9622 10894: contig of 1273 bp in length  
\* 10895 10994: gap of 100 bp  
\* 10995 12113: contig of 1119 bp in length  
\* 12114 12213: gap of 100 bp  
\* 12214 14077: contig of 1864 bp in length  
\* 14078 14177: gap of 100 bp  
\* 14178 15552: contig of 1375 bp in length  
\* 15553 15652: gap of 100 bp  
\* 15653 16845: contig of 1193 bp in length  
\* 16846 16945: gap of 100 bp  
\* 16946 18701: contig of 1756 bp in length  
\* 18702 18801: gap of 100 bp  
\* 18802 20280: contig of 1479 bp in length  
\* 20281 20380: gap of 100 bp  
\* 20381 22043: contig of 1663 bp in length  
\* 22044 22143: gap of 100 bp  
\* 22144 23464: contig of 1321 bp in length  
\* 23465 23564: gap of 100 bp  
\* 23565 25268: contig of 1704 bp in length  
\* 25269 25368: gap of 100 bp  
\* 25369 26869: contig of 1501 bp in length  
\* 26870 26969: gap of 100 bp  
\* 26970 28961: contig of 1992 bp in length  
\* 28962 29061: gap of 100 bp  
\* 29062 31160: contig of 2099 bp in length  
\* 31161 31260: gap of 100 bp  
\* 31261 32989: contig of 1729 bp in length  
\* 32990 33089: gap of 100 bp

\* 33090 35728: contig of 2639 bp in length  
\* 35729 35828: gap of 100 bp  
\* 35829 37664: contig of 1836 bp in length  
\* 37665 37764: gap of 100 bp  
\* 37765 39295: contig of 1531 bp in length  
\* 39296 39395: gap of 100 bp  
\* 39396 41939: contig of 2544 bp in length  
\* 41940 42039: gap of 100 bp  
\* 42040 44376: contig of 2337 bp in length  
\* 44377 44476: gap of 100 bp  
\* 44477 47182: contig of 2706 bp in length  
\* 47183 47282: gap of 100 bp  
\* 47283 50373: contig of 3091 bp in length  
\* 50374 50473: gap of 100 bp  
\* 50474 54496: contig of 4023 bp in length  
\* 54497 54596: gap of 100 bp  
\* 54597 57825: contig of 3229 bp in length  
\* 57826 57925: gap of 100 bp  
\* 57926 64181: contig of 6256 bp in length  
\* 64182 64281: gap of 100 bp  
\* 64282 70174: contig of 5893 bp in length  
\* 70175 70274: gap of 100 bp  
\* 70275 74993: contig of 4719 bp in length  
\* 74994 75093: gap of 100 bp  
\* 75094 80432: contig of 5339 bp in length  
\* 80433 80532: gap of 100 bp  
\* 80533 85565: contig of 5033 bp in length  
\* 85566 85665: gap of 100 bp  
\* 85666 95772: contig of 10107 bp in length  
\* 95773 95872: gap of 100 bp  
\* 95873 108321: contig of 12449 bp in length  
\* 108322 108421: gap of 100 bp  
\* 108422 118802: contig of 10381 bp in length  
\* 118803 118902: gap of 100 bp  
\* 118903 130159: contig of 11257 bp in length  
\* 130160 130259: gap of 100 bp  
\* 130260 146057: contig of 15798 bp in length  
\* 146058 146157: gap of 100 bp  
\* 146158 167429: contig of 21272 bp in length.

FEATURES  
source  
1. .167429  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .1114  
/note="assembly\_fragment"  
1215. .2323  
/note="assembly\_fragment"  
2424. .3499  
/note="assembly\_fragment"  
3600. .4691  
/note="assembly\_fragment"  
4792. .6613  
/note="assembly\_fragment"  
6714. .8003  
/note="assembly\_fragment"  
8104. .9521  
/note="assembly\_fragment"  
9622. .10894  
/note="assembly\_fragment"  
10995. .12113  
/note="assembly\_fragment"  
12214. .14077  
/note="assembly\_fragment"  
14178. .15552  
/note="assembly\_fragment"  
15653. .16845  
/note="assembly\_fragment"  
16946. .18701  
/note="assembly\_fragment"  
18802. .20280  
/note="assembly\_fragment"

misc\_feature  
1. .1114  
misc\_feature  
1215. .2323  
misc\_feature  
2424. .3499  
misc\_feature  
3600. .4691  
misc\_feature  
4792. .6613  
misc\_feature  
6714. .8003  
misc\_feature  
8104. .9521  
misc\_feature  
9622. .10894  
misc\_feature  
10995. .12113  
misc\_feature  
12214. .14077  
misc\_feature  
14178. .15552  
misc\_feature  
15653. .16845  
misc\_feature  
16946. .18701  
misc\_feature  
18802. .20280



**TITLE** Direct Submission  
**JOURNAL** Submitted (14-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
**COMMENT** On Feb 25, 2000 this sequence version replaced gi:6693396.  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc/Index.shtml>  
 Project Information -----  
 Center project name: H.NH0744D14  
 Summary Statistics -----  
 Sequencing vector: M13; 99%  
 Chemistry: Dye-terminator; plasmid; 1%  
 Chemistry: Dye-terminator ET; 99% of reads  
 Chemistry: Dye-terminator Big Dye; 1% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 197586 bases at least Q40  
 Consensus quality: 198226 bases at least Q30  
 Consensus quality: 198612 bases at least Q20  
 Insert size: 205000; agarose-fp  
 Insert size: 199737; sum-of-contigs  
 Quality coverage: 6.75 in Q20 bases; agarose-fp  
 Quality coverage: 6.95 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 3193: contig of 3193 bp in length  
 \* 3194 3293: gap of unknown length  
 \* 3294 17500: contig of 14207 bp in length  
 \* 17501 17600: gap of unknown length  
 \* 17601 33944: contig of 16344 bp in length  
 \* 33945 34044: gap of unknown length  
 \* 34045 52735: contig of 18691 bp in length  
 \* 52736 52835: gap of unknown length  
 \* 52836 75483: contig of 22648 bp in length  
 \* 75484 75583: gap of unknown length  
 \* 75584 117631: contig of 42048 bp in length  
 \* 117632 117731: gap of unknown length  
 \* 117732 200337: contig of 82606 bp in length.  
**FEATURES** Location/Qualifiers  
 source 1..200337  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7L"  
 /clone="RP11-744D14"  
 BASE COUNT 64278 a 36060 c 36848 g 62535 t 616 others  
 ORIGIN  
 Query Match 2.9% Score 41.2; DB 61; Length 200337;  
 Best Local Similarity 49.1%; Pred.No.4.9; Indels 0; Gaps 0;  
 Matches 109; Conservative 0; Mismatches 113;  
 QY 1065 ttttcaatcttccatgctcacatctgatgtg99cgtgtgcaattaggacattcata 1124  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 79066 TTTTITTTTTCGTCAAAATGATGTCGTTCCTGTATATCTTGATATTAGTCTTGCCA 79125  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1125 tacaacaggagggtgattggttatactatttggccttttctgctgttgggaata 1184  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 79126 TGTAAATAGTTTGCAAAATATTTTCTCCCATTTTGACAGATTGCTCCACTCTGTTGATGGA 79185  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1185 ttcatatatcaaaaacagagaagatcctcaatacagagcccttgagaatggaac 1244  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 79186 ATGGAATATTTTGAAGAGAAATTAAGCGAGCTATTCACGAAAGTGCCCTTCAAAATATCAAC 79245  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1245 ttgatcatgaatatgctctccttggagatcaaatgacaca 1286



Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS  
1 (bases 1 to 22696)  
Adams, M. and Venter, J.C.

**TITLE** Direct Submission  
**JOURNAL** Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT  
This sequence was identified as CDM:10211247 by the submitter.  
For further information on this sequence you may e-mail to  
fly@celera.com.

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

**FEATURES**

1.22696  
source

1. 22696

1. : 22050  
/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT	7343 a	4511 c	3971 q	6871 t
/uu_xref= caxon: /zz/				

## ORIGIN

```
Query Match      2.88; Score 40.8; DB 42; Length 22696;
Best Local Similarity 50.0%; Pred. No. 6.9;
Matches 102; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
```

Ov 1234 agaatgaaaacttggatcatggaatatagtctccttgaaggatcaagatgacacacccaagaa 1293

[illegible]

Ov 1294 aaccatagctagctttgcccaacaatcttttcatcaacagtttttaataatttcgtagggg 1353

[illegible]

Ov 1354 qagagagatcgaagatactaatctaattaatcgaacgtctattatataagttcgaagattttttgattta 1413

C7	1334	gugagugacgcgagatcaaaatttctttgtccatcatataggcggagggccccccc	
		{                               }	
D6	15307	cacatatattattgtgaaccaatcttcgactcttttttttgtaattatttttaagtgtttt	

Qy 1414 tttatttttgaataaaaaaa 1437

[illegible]

Search completed: August 1, 2000, 13:36:13  
Job time: 57635 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: August 1, 2000, 19:50:24 ; Search time 47.63 Seconds  
(without alignments)  
166.593 Million cell updates/sec

Title: US-09-147-955-10

Perfect score: 1723

Sequence: 1 MSSSSRRWRNEGMRRTLL.....GNLDHEYSLLEDQDDTPRKP 335

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1723	100.0	335	1 W92951	W09905287 Seq ID 1
2	106	6.2	402	1 W13497	Plant plastid phos
3	105	6.1	1030	1 W29682	Human thiazide-sen
4	102	5.9	336	1 W27294	Human J59-41 sece
5	98	5.7	375	1 W20731	H. pylori inner me
6	97	5.6	679	1 R10848	Human gibbon ape l
7	97	5.6	679	1 R72969	GALV receptor prot
8	97	5.6	679	1 W70498	Human sodium-lithi
9	97	5.6	679	1 W96998	The gibbon ape leu
10	96.5	5.6	462	1 W97820	Staphylococcus aur
11	95.5	5.5	456	1 P91895	Protein sequence o
12	95	5.5	383	1 W98461	H. pylori GHPO 718
13	94	5.4	457	1 W42430	Escherichia coli g
14	93.5	5.4	312	1 Y11033	H. pylori ORF 04ep
15	93.5	5.4	407	1 Y11034	H. pylori ORF 04ep
16	92	5.3	365	1 R42280	PGF2 receptor codi
17	92	5.3	474	1 W58862	T. halophilus xylul
18	91	5.3	361	1 R42281	PGF2 receptor codi
19	90.5	5.3	415	1 R05539	Rat D2 dopamine re
20	90.5	5.3	415	1 W09388	Mouse dopamine D2
21	90.5	5.3	415	1 Y01598	Rat D2 dopamine re
22	90.5	5.3	444	1 R10544	D2 dopamine recept
23	90.5	5.3	444	1 W09387	Mouse dopamine D2
24	90	5.2	446	1 W88325	Flippase involved
25	88.5	5.1	432	1 W14081	S. thermophilus exo
26	88.5	5.1	467	1 W05741	Presenilin-1 I14
27	88.5	5.1	473	1 W22180	S. thermophilus exo
28	87.5	5.1	401	1 W92940	Pea Glucose-6-phos
29	87.5	5.1	506	1 W81978	Ehrlichia sp. E74.
30	87.5	5.1	3033	1 R33538	NANBH virus strain
31	86	5.0	365	1 W48086	Human dendritic ce
32	85.5	5.0	348	1 R77224	Hamster ileal/rene
33	85.5	5.0	383	1 W01664	p(rat-edg), G-prot

## ALIGNMENTS

RESULT 1

W92951

ID W92951 standard; Protein; 335 AA.

AC W92951:

DT 14-MAY-1999 (first entry)

DE W09905287 Seq ID 11.

KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.

OS Perilla frutescens.

PN W09905287-A1.

PD 04-FEB-1999.

PF 16-JUL-1998; J03199.

PR 25-JUL-1997; JP-200571.

PA (SUNR ) SUNTORY LTD.

PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,

PI Yamazaki M;

DR WPI: 99-142940/12.

DR N-PSDB: X02830.

PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be

PT used to transform plants for improvement of plant coloration

PS Claim 2; Page 81-82; 89pp; Japanese.

CC This invention describes the isolation of plant proteins which have

CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used

CC in the production of plant varieties with improved colour, for example in

CC the production of flower varieties with desired coloration and cut

CC flowers harvested from them.

SQ Sequence 335 AA;

Query Match

Best Local Similarity 100.0%; Score 1723; DB 1; Length 335;

Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSSSSRRWRNEGMRRTLLGLGLQVLSFDLAIMTFASLSVSTTVDDAPLMTSFYTVV 60

Db 1 MSSSSRRWRNEGMRRTLLGLGLQVLSFDLAIMTFASLSVSTTVDDAPLMTSFYTVV 60

Qy 61 ALLYGTILLYRRHFLVPWYVALLGFDVHGNYLVNKAFLTSITSVSLDCWTVWSI 120

Db 61 ALLYGTILLYRRHFLVPWYVALLGFDVHGNYLVNKAFLTSITSVSLDCWTVWSI 120

Qy 121 IFTWMFLGTYKVQVFGAAICVGGLLLVLLSDSGVTAAGSNPLLGDFLVITGSLFTLS 180

Db 121 IFTWMFLGTYKVQVFGAAICVGGLLLVLLSDSGVTAAGSNPLLGDFLVITGSLFTLS 180

Qy 181 TVGQYCVKRDRIEVMAMGVFGMLISATEITVLERALNSMQHSTGLLAAYVYVALSS 240

Db 181 TVGQYCVKRDRIEVMAMGVFGMLISATEITVLERALNSMQHSTGLLAAYVYVALSS 240

Qy 241 FLFCFLTLPFLKMSGAFFNLISMLTSDMWAVAIRFIYNOEVDWLYLAFCLVWVGIFY 300

Db 241 FLFCFLTLPFLKMSGAFFNLISMLTSDMWAVAIRFIYNOEVDWLYLAFCLVWVGIFY 300

Qy 301 TKTEKDPNPNTRALENGNDHEYSLEDDQDDTPRKP 335

Db 301 TKTEKDPNPNTRALENGNDHEYSLEDDQDDTPRKP 335

	RESULT	3	
QY	W29682		
110	ID	W29682 standard; Protein; 1030 AA.	
Db	AC	W29682:	
QY	DT	21-DEC-1998 (first entry)	
110	DE	Human thiazide-sensitive NaCl cotransporter TSC.	
Db	KW	Thiazide-sensitive Na-Cl cotransporter; TSC; human; ion transport;	
QY	KW	Gitelman's syndrome; Bartter's syndrome; hypokalaemic alkalosis;	
110	KW	hypocalciuria; hypomagnesaemia; diagnosis; therapy.	
Os	OS	Homo sapiens.	
110	FH	Key	
Db	FH	Location/Qualifiers	

```

QY 258 FFNLSMLTSDMMAVAIRFTIYNQEVDM 284
DB 317 FFS---YRADI-----FVQNLVDPW 333

RESULT 4
W27294
ID W27294 standard; Protein; 336 AA.
AC W27294;
DE 19-MAY-1998 (first entry)
KW Human J59-41 secreted protein.
KW Secreted protein; anti-inflammatory; immune stimulant; suppressant;
KW human; cytokine; autoimmune disease; regulator; activin; inhibin;
KW J59-41.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..184
FT Protein /label= signal
FT 185..336
FT /label= J59-41
FT /note= "secreted protein"
FN W09740151-A2.
PD 30-OCT-1997.
PF 14-APR-1997; U06125.
PR 10-JAN-1997; US-781226.
PR 19-APR-1996; US-635311.
PA (GEMV ) GENETICS INST INC.
PI Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA,
PI Spaulding V;
DR WPI: 97-535835/49.
DR N-PSDB: T91305.
PT Nucleic acids encoding novel secreted proteins - useful e.g. as an
PT anti-inflammatory, immune stimulant or suppressor, etc.
PS Claim 21; Page 63-64; 81pp; English.
CC This sequence represents a novel secreted protein, J59-41, which is
CC isolated from a clone, J59-41 (ATCC 98028), derived from a human PBMC
CC cDNA library. The nucleic acid and encoded J59-41 secreted protein can be
CC used for research purposes (as markers for tissues, molecular weight
CC markers for gels, primers, probes, etc.), for nutrition (as C, N or
CC carbohydrate source), as a cytokine for cell proliferation and
CC differentiation activity, as immune stimulants or suppressors e.g. for
CC viral, bacterial or fungal infections, for autoimmune diseases such as
CC multiple sclerosis or systemic lupus erythematosus, to regulate
CC haematopoiesis, for tissue growth, as an activin or inhibin, or having
CC chemotactic, chemokinetic, haemostatic and thrombolytic, receptor/ligand,
CC anti-inflammatory or tumour inhibitory activities.
SQ Sequence 336 AA;

Query Match 5.9%; Score 102; DB 1; Length 336;
Best Local Similarity 22.3%; Pred. No. 0.0017;
Matches 62; Conservative 41; Mismatches 91; Indels 84; Gaps 13;

QY 52 MSFTYTVALLYGTLILRRHKLFLVPWYVALLGFDVHGNYLVNKAFLTSTVSIL 111
DB 86 VQFAPYSIFGLIELQIQDKRR--IPGKTYMIIAFLTVGTGMLGSLNTSLGYLNYPTQVIF 143
QY 112 DCITVNSIIFTWFLTKYSYQFVGNAICVG-GLLVLLSDSGVTAAGSN-----PL 164
DB 144 KCCKLIPVMLGGVFIOGKRYNVAD-VSAAICMSGLIWFITLADS--TTAPNPNRLVLSYS 201
QY 165 LGDFLVITGSLFTLSTVGQ--EYCVKRKDRIE-----VVMAGIVGEML 206
DB 202 IGFVILIG--LCTSGIGPANTCAKNPRTYGYAFLSFTGYFGISFVIALIKIFGAL 259
QY 207 ISATEITVLERNALSSMOWSTGLAAAYVVALSYVALLSFLFCTLTLPFLKXSGAAFFNLMLTS 266
DB 260 IAVTV-----TTGRAMTIVL--SFIFA-KPTFTQ----- 287
QY 267 DMWAVAIRFTIYNQEVDMWYLAFLCLVVGIF--IYTK 302
DB 288 -----YWSGLLVVLGIFLWYSK 306

```

```

RESULT 5
W20731
ID W20731 standard; protein; 375 AA.
AC W20731;
DE 16-JUL-1997 (first entry)
KW H. pylori inner membrane protein, 06cp11118orf6.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.
PN W09640893-A3.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PF 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI: 97-052306/05.
DR N-PSDB: T67984.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 56; Page 1146-1147; 1481pp; English.
CC The present sequence is a Helicobacter pylori inner membrane protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 375 AA;

```

```

Query Match 5.7%; Score 98; DB 1; Length 375;
Best Local Similarity 19.9%; Pred. No. 0.0057;
Matches 65; Conservative 57; Mismatches 106; Indels 98; Gaps 18;

QY 52 MSFTYTVVAL--LYGTILLYR-RHKFLVPWYV-YALLGFDVHG-----NYLVNK----- 98
DB 72 LSIPTFYVADMGSYVHSSISMYSGSNDALIQMWGHNVAFAVFTSGVIGTIYIFLPKESGQ 131
QY 99 ---AFELTSTSVSILDCWTIWSIIFTWFLTKYSVY----OFVGAACVGGLLVLL 151
DB 132 PIFYKTLFS-----FWSLMFYIWAAGHHLIYSTVPDWQTLSSVFSVLL- 180
QY 152 SDSGVTAAGSNPLGDLFLVITGSILFTLSTVGYQCYVKKRDRIEVAMIGVGM-ISA 209
DB 181 -----PSNGT-----AINMLLTMRGQWHLKESPLIKFLVLASTFYMLSTLEG 223
QY 210 TEITVLERNALSSM-QWSTGL----LAAYVVALSSFLFCTLTLPFLK---MSGAAF-FN 260
DB 224 SIQAKSVNALAHFTDWIIIGHVHDGVLGVTGFTLIASMY-HMTPLRKREIYSGRLVDQ 282
QY 261 LSLMT-----SDMWAVAIRFTIYNQEVDM-----WLY 286
DB 283 FWIMTLGLIVLFSSMIAGITQGMHWRDQYGNLTQYQIDITVKALIPYINIRGVGLMY 342
QY 287 YLAFCLVVGIFIV---TKTEKDPN 308
DB 343 FIGIIFAYNFMITAGKKLEREPN 368

RESULT 6
R10848
ID R10848 standard; Protein; 679 AA.

```



CC Including manic depressives, to treatment with lithium salts.  
 CC Probes and primers for PIT-1, PIT-2 (see W70499) and BNPI (see  
 CC W70500) can be used in diagnostic tests useful for genetic  
 CC screenings to predict whether a patient will respond to lithium  
 CC treatment. The test is also a screen for susceptibility to, and  
 CC extent of, manic depressive illness, and is suitable for screening  
 CC newborns.  
 CC Sequence 679 AA;

Query Match 5.6%; Score 97; DB 1; Length 679;  
 Best Local Similarity 21.7%; Pred. No. 0.018;  
 Matches 66; Conservative 49; Mismatches 105; Indels 84; Gaps 16;

QY 76 LVPWYVALLGFEVDHGNLYLVNKAFLT-SITSVSLDCW-TIVWSIIFTWMLGTRYSV 133  
 DB 18 LVDYLMMLILGFI-----TAFVLAFSVGANDVANSGTAVGSGVVTLKQACILASI 68  
 QY 134 YQFVGAACVIG-----GLLLVLL--SDSGVTAAGS-----NPLL 165  
 DB 69 FEIVGSVLLGAKYSETIRKGLIDVEMYNSTQGLLMAGSVSAMFGSAVWQVASFLLKLPIS 128  
 QY 166 GDFLVTGSLTSTVQGYCVKRDRIEV-----AMIGVF-GMLISATEITVLER 217  
 DB 129 GTHCIVGATIGFSLVAKGQE-GVKWSELIKIVMSFWVSPLLSGIMSGILFFLVRAFLHK 187  
 QY 218 NALSSMOWSTGLLAAYVYVYALSSELFCT--LTPFLKMSGAAPFNLSMLTSDMWAVAIR 275  
 DB 188 ----ADVPNGRLALPVFYA-----CTVGINLFSIMYTGAPLLGFDKL--PLWGTILIS 235  
 QY 276 FIYNQEDWLYLAFCLVGVGIFITYTTEK-----DPNTRALENGNLDHEYSLEL 327  
 DB 236 -----VCAVFCALIVWFFVCPMKRKIEREIKSPSESLMEKKN-----SLKED 281  
 QY 328 QDDT 331  
 DB 282 HEET 285

## RESULT 9

W96998  
 ID W96998 standard; Protein; 679 AA.  
 AC W96998;  
 DT 23-APR-1999 (first entry)  
 DE The gibbon ape leukemia virus (GALV) receptor.  
 KW Human gibbon ape leukemia virus receptor; GALV; substrate transport;  
 KW ion transport; immune regulatory activity.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Modified\_site 100 /note= "potential N-glycosylation site"  
 FT Misc\_difference 295 /note= "encoded by AAC"  
 FT Misc\_difference 363 /note= "encoded by ACT"  
 FT Modified\_site 374 /note= "potential N-glycosylation site"  
 FT Modified\_site 418 /note= "potential N-glycosylation site"  
 FT Modified\_site 497 /note= "potential N-glycosylation site"  
 FT Region 743..775 /note= "repeat region"  
 FT US5874264-A.  
 PN 23-FEB-1999.  
 PF 08-MAY-1995; 436900.  
 PR 25-MAR-1991; US-674287.  
 PR 24-AUG-1989; US-398351.  
 PR 08-MAY-1995; US-436900.  
 PA (AMCY ) AMERICAN CYANAMID CO.  
 PI Ohara BM;  
 DR WPI; 99-179981/15.  
 DR N-PSDB; X15171.

PT Recombinant gibbon ape leukemia virus receptor proteins - and  
 PT plasmids containing nucleotide sequences encoding such proteins  
 PS Claim 1; Fig 7A-B; 34pp; English.  
 CC The present sequence represents a human gibbon ape leukaemia virus  
 CC (GALV) receptor protein. The cDNA sequence was isolated from HL60  
 CC cells. The GALV receptor proteins are useful in experimental  
 CC manipulation of the GALV host, in the analysis of virus/receptor  
 CC interactions, and in the elucidation of the normal role of the  
 CC receptor, which is thought to include substrate/ion transport and  
 CC immune regulatory activity.  
 SQ Sequence 679 AA;

Query Match 5.6%; Score 97; DB 1; Length 679;  
 Best Local Similarity 21.7%; Pred. No. 0.018;  
 Matches 66; Conservative 49; Mismatches 105; Indels 84; Gaps 16;

QY 76 LVPWYVALLGFEVDHGNLYLVNKAFLT-SITSVSLDCW-TIVWSIIFTWMLGTRYSV 133  
 DB 18 LVDYLMMLILGFI-----TAFVLAFSVGANDVANSGTAVGSGVVTLKQACILASI 68  
 QY 134 YQFVGAACVIG-----GLLLVLL--SDSGVTAAGS-----NPLL 165  
 DB 69 FEIVGSVLLGAKYSETIRKGLIDVEMYNSTQGLLMAGSVSAMFGSAVWQVASFLLKLPIS 128  
 QY 166 GDFLVTGSLTSTVQGYCVKRDRIEV-----AMIGVF-GMLISATEITVLER 217  
 DB 129 GTHCIVGATIGFSLVAKGQE-GVKWSELIKIVMSFWVSPLLSGIMSGILFFLVRAFLHK 187  
 QY 218 NALSSMOWSTGLLAAYVYVYALSSELFCT--LTPFLKMSGAAPFNLSMLTSDMWAVAIR 275  
 DB 188 ----ADVPNGRLALPVFYA-----CTVGINLFSIMYTGAPLLGFDKL--PLWGTILIS 235  
 QY 276 FIYNQEDWLYLAFCLVGVGIFITYTTEK-----DPNTRALENGNLDHEYSLEL 327  
 DB 236 -----VCAVFCALIVWFFVCPMKRKIEREIKSPSESLMEKKN-----SLKED 281  
 QY 328 QDDT 331  
 DB 282 HEET 285

## RESULT 10

W97820  
 ID W97820 standard; Protein; 462 AA.  
 AC W97820;  
 DT 21-MAY-1999 (first entry)  
 DE Staphylococcus aureus tetracycline resistance protein.  
 KW Tetracycline resistance; antibiotic; antibacterial; screening;  
 KW infection; otitis media; endocarditis; abscess; conjunctivitis;  
 KW epididymitis; toxic shock syndrome; sepsis; septic arthritis;  
 KW diagnosis; therapy; vaccine.  
 OS Staphylococcus aureus.  
 PN EP-893499-A2.  
 PD 27-JAN-1999.  
 PF 10-JUL-1998; 305495.  
 PR 23-JUL-1997; US-898976.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Burnham MKR, Lonetto MA, Warren PV;  
 DR WPI; 99-097783/09.  
 DR N-PSDB; X07357.  
 PT New Staphylococcus aureus tetracycline resistant protein and gene -  
 PT useful as diagnostic reagents and for prevention and treatment of  
 PT staphylococcal infections, including toxic shock syndrome and otitis  
 PT media  
 PS Claim 12; Page 27-29; 35pp; English.  
 CC This is the amino acid of a novel tetracycline resistance protein  
 CC (TR) of Staphylococcus aureus WCUH29 (NCIMB 40771) that is involved  
 CC in sugar or drug transport across bacterial cytoplasmic membranes.  
 CC TR polynucleotide (see X07357) has been isolated from an S. aureus  
 CC DNA library. Vectors comprising the claimed polynucleotide, host  
 CC cells and a process for producing the TR protein are claimed. TR



(HUMA-) HUMAN GENOME SCI INC.  
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 PI Al-Garawi A, Kleianthous H, Miller C, Oomen RP, Tomb J;  
 DR WPI: 98-542293/46.  
 DR N-PSDB; X14180.  
 PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases  
 PS Claim 8: Page 815-817; 2054pp; English.  
 CC This sequence represents a Helicobacter pylori GHPO protein of the  
 CC invention. The polypeptides can be used for preventing or treating  
 CC Helicobacter infections, and gastroduodenal diseases associated with  
 CC these infections, including acute, chronic, and atrophic gastritis, and  
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
 CC used for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.  
 SQ Sequence 383 AA;

Query Match 5.5%; Score 95; DB 1; Length 383;  
 Best Local Similarity 20.6%; Pred. No. 0.013;  
 Matches 54; Conservative 51; Mismatches 79; Indels 78; Gaps 14;  
 QY 58 TVVALLGTYLLYRRHRKFLVPMWYALLGFDVHGNYLVNKAFLTSTTSVSILDCWTV 117  
 DB 30 TVVEILFGSVGAY-----VGPIE-----PTKFEIMSEI----- 58  
 QY 118 WSIIFTWMLGTYKSYVQF--VGAIC--VGLLLVLLSDSGVTAAGSN--PLLGDFLVI 171  
 DB 59 -GFLFLMFLGCGVEIYLFKRLGVSLLKRIPAYLLIYLTSLFILTSLNLEPI---EMVI 114  
 QY 172 TGS1-LFTLSVNGGEYCVKRD--RIEVMAMIGVFGMLISATEITVLE---RNALSSMQ 224  
 DB 115 FPISLGMWTLVDY---RKEILWLDVLKVGIVGELSLIFGVVGVYSHGLGMDLI 171  
 QY 225 WSTGLLAAYVYVALSSF-LFCTLPFLKMGSAFFNLMLTSDMVAIVRTFI----- 277  
 DB 172 KDLGILIVFLIIIVAFQIKTLF-----WFRPHLKLFLVMPKSSQ 211  
 QY 278 YNQEDWLYIYAFCLVVGIF 299  
 DB 212 FNQDVFSLMLFFSLVAIVVWL 233

RESULT 13  
 W42430  
 ID W42430 standard; Protein; 457 AA.  
 AC W42430;  
 DT 06-JUL-1998 (first entry)  
 DE Escherichia coli gusB gene product.  
 KW GusB gene; glucuronide repressor; GusR; transgene detection; assay.  
 OS Escherichia coli.  
 PN W09749813-A2.  
 PD 31-DEC-1997.  
 PR 26-JUN-1997; U11427.  
 PR 26-JUN-1996; US-020621.  
 PA (CAMP-) CAMBIA BIOSYSTEMS LLC.  
 PI Jefferson RA, Leader M, Wilson KJ;  
 DR WPI: 98-077176/07.  
 DR N-PSDB; W03430.  
 PT Glucuronide repressor gene - useful to develop products to detect  
 PT glucuronide(s), control gene expression or identify glucuronide  
 PT transport proteins  
 PS Disclosure; Page 47-48; 84pp; English.  
 CC This polypeptide comprises the predicted gene product of the gusB  
 CC gene of the gusABC operon (see W03430) of Escherichia coli.  
 CC The invention relates the the glucuronide repressor gene (see  
 CC W03424) located upstream of the gusA gene in the gus operon.  
 CC The repressor acts by binding to gusA operator sequences (see also  
 CC W03429), thus preventing transcription, this repression being  
 CC relieved when a glucuronide substrate binds to the repressor and  
 CC inactivates it. The repressor is used in claimed methods for  
 CC controlling expression of a transgene, detecting glucuronides in a

CC sample and isolating glucuronides from a sample.  
 SQ Sequence 457 AA;  
 Query Match 5.5%; Score 94; DB 1; Length 457;  
 Best Local Similarity 18.8%; Pred. No. 0.022;  
 Matches 82; Conservative 71; Mismatches 125; Indels 158; Gaps 18;  
 QY 17 RTLLGLGLGQ-----LVSEDLAIMTFSASLVSSTVDAPLTMTSTTTTVV--- 60  
 DB 8 RTIVGYSGLGVANNFAMGALFLLSYTVDVAGVGAAGAATML--LLVRVDFADAFVAG 65  
 QY 61 -----ALLYGT--ILLYRRHKFLVPMWY-----YALLGFDVHGNY-L 95  
 DB 66 RVVDSVNRWKGKFPFLFGTAPLMIFSVLVFWVPTDWSHGSKVVYAYLYTMGLGCLYS 125  
 QY 96 VNKAFF--LTSITS-----VSILDCW 114  
 DB 126 VNIPYGLATAMTQQPQSRARLGAARGTAASLTFVCLAFLLGIPSIKSSPPEMVSVYHFW 185  
 QY 115 TIWWSI-----IFTWMLGTYKSYVQFVGA-----AICVGLLLVLLSD 153  
 DB 186 TIVLAIAGMWLYFTCFKSTRENVYRIVAQPSLNISLQTLKRNRLFLMCLIGA-LCVLIST 244  
 QY 154 SGVTAAG-----SNPLLGDELIVITGSLFTLSTV---GOEYCVKRDRIEVMAMIG 201  
 DB 245 FVNSASSLFVRYVINDTGLFTVLVLVQNLVGTVASAPLVPGMVARGKNTFLIGALLG 304  
 QY 202 VFGML-----ISATEITVLERNALSSMOWSTGLLAAVYVY 236  
 DB 305 TCGYLLFFWVSWSLPVVALVAIAIASIGQGVMTVMNALEADTVEYGEYLTGVRIEGLTY 364  
 QY 237 ALSFP-----LCTLTLPFLKMGSAFFNLMLTSDMVAIVRTFIYNOEDWLYLA 289  
 DB 365 SLFSPKCGQAICGSPAFILGLSG---YIANQVQTPVEVIMGIRTSIALVPCGFM-LLA 420  
 QY 290 FCLVVVGIFITKTEK 305  
 DB 421 FVI-----IWFPLTDK 432  
 RESULT 14  
 Y11033  
 ID Y11033 standard; Protein; 312 AA.  
 AC Y11033;  
 DT 08-JUN-1999 (first entry)  
 DE H. pylori ORF 04ep41903\_16667055\_cl\_37 inner membrane protein.  
 KW vaccine; probe; diagnostic; ORF; cell envelope protein;  
 KW secreted protein; cytoplasmic protein; cellular protein.  
 OS Helicobacter pylori.  
 PN W09824475-A1.  
 PD 11-JUN-1998.  
 PR 05-DEC-1997; U22104.  
 PR 14-JUL-1997; US-891928.  
 PR 05-DEC-1996; US-759625.  
 PR 25-MAR-1997; US-823745.  
 PA (ASTR) ASTRA AB.  
 PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;  
 DR WPI: 98-333051/29.  
 DR N-PSDB; X30562.  
 PT New isolated Helicobacter pylori nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of infection by  
 PT H. pylori and other Helicobacter species  
 PS Claims 37, 41; Page 197; 339pp; English.  
 CC Recombinant or substantially pure preparations of H. pylori polypeptides  
 CC are disclosed, together with the nucleic acids encoding them. In all,  
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,  
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.  
 CC Vaccines containing the nucleic acids or proteins are claimed, as are  
 CC probes containing at least 8 nucleotides from the nucleic acid  
 CC sequences. The vaccines are useful for treating or reducing the risk of  
 CC H. pylori infections, and the probes can be used diagnostically for  
 CC detecting the presence of Helicobacter in a sample. The products are

CC also of use in screening for compounds having the ability to interfere  
CC with the H. pylori life cycle or to inhibit H. pylori infection.  
SQ Sequence 312 AA;

Query Match 5.4%; Score 93.5; DB 1; Length 312;  
Best Local Similarity 20.2%; Pred. No. 0.014;  
Matches 54; Conservative 46; Mismatches 112; Indels 55; Gaps 10;  
QY 18 TLGLGLGOLVSFDLAIMTFASLSSTVDA-PLTMSFTTYTVVALLYTILLYRR---- 72  
DB 52 TTGLPIFGSLIFSATKTSNLSLIDKLADAKSVQMPYLGAVFSLLLALVYLLKLPDV 111  
QY 73 -----HKFLVPWYVWYALLGFDVHGNYLVNKAFELTITSVSILDCWTIV 117  
DB 112 EKEMPKETQKSLFSHKHFV-----FGALGIFVYVGEVAIGSFVLVSFEKLLNLDQSSA 167  
QY 118 WSIIFTW-----MFLGT-----KYSVYQFVGAAICVGLLIVLLSDSGVTAAGSN 162  
DB 168 HYLIVYWGAMVGRFLGSALMNKIAPNKYLAFNALSSIILIALAILI-----GGKIALFAL 223  
QY 163 PLLGDFLVITGSLITFLSTVGOEYCVKRDRIEVVAMIGVFGMLISATE-----ITVLE 216  
DB 224 TVGFGFNSIMFTIETSLATNLGHLTSKASGVISMAIVG--GALIPPIQGVVTDMLTATE 281  
QY 217 RNALSSMOWSTGLLAAYV--YALSSF 241  
DB 282 SNLLYA--YSVPLLCYFYILFFALKGY 306

## RESULT 15

Y11034  
ID Y11034 standard; Protein; 407 AA.  
AC Y11034;  
DT 08-JUN-1999 (first entry)  
DE H. pylori ORF 04ep41903\_19689182\_cl\_43 inner membrane protein.  
KW Vaccine; probe; diagnostic; ORF; cell envelope protein;  
OS secreted protein; cytoplasmic protein; cellular protein.  
OS Helicobacter pylori.  
PN W09824475-A1.  
PD 11-JUN-1998.  
PF 05-DEC-1997; U22104.  
PR 14-JUL-1997; US-891928.  
PR 05-DEC-1996; US-759625.  
PR 25-MAR-1997; US-823745.  
PA (ASTR ) ASTRA AB  
PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;  
DR WPI; 98-333051/29.  
DR N-PSDB; X30563.  
PT New isolated Helicobacter pylori nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of infection by  
PT H. pylori and other Helicobacter species  
PS Claims 37, 41; Page 198-199; 339pp; English.  
CC Recombinant or substantially pure preparations of H. pylori polypeptides  
CC are disclosed, together with the nucleic acids encoding them. In all,  
CC 97 ORFs are shown. The proteins are variously cell envelope proteins,  
CC cytoplasmic proteins, secreted proteins or other cellular proteins.  
CC vaccines containing the nucleic acids or proteins are claimed, as are  
CC probes containing at least 8 nucleotides from the nucleic acid  
CC sequences. The vaccines are useful for treating or reducing the risk of  
CC H. pylori infections, and the probes can be used diagnostically for  
CC detecting the presence of Helicobacter in a sample. The products are  
CC also of use in screening for compounds having the ability to interfere  
CC with the H. pylori life cycle or to inhibit H. pylori infection.  
SQ Sequence 407 AA;

Query Match 5.4%; Score 93.5; DB 1; Length 407;  
Best Local Similarity 20.2%; Pred. No. 0.021;  
Matches 54; Conservative 46; Mismatches 112; Indels 55; Gaps 10;  
QY 18 TLGLGLGOLVSFDLAIMTFASLSSTVDA-PLTMSFTTYTVVALLYTILLYRR---- 72  
DB 52 TTGLPIFGSLIFSATKTSNLSLIDKLADAKSVQMPYLGAVFSLLLALVYLLKLPDV 111  
QY 73 -----HKFLVPWYVWYALLGFDVHGNYLVNKAFELTITSVSILDCWTIV 117  
DB 112 EKEMPKETQKSLFSHKHFV-----FGALGIFVYVGEVAIGSFVLVSFEKLLNLDQSSA 167  
QY 118 WSIIFTW-----MFLGT-----KYSVYQFVGAAICVGLLIVLLSDSGVTAAGSN 162  
DB 168 HYLIVYWGAMVGRFLGSALMNKIAPNKYLAFNALSSIILIALAILI-----GGKIALFAL 223  
QY 163 PLLGDFLVITGSLITFLSTVGOEYCVKRDRIEVVAMIGVFGMLISATE-----ITVLE 216  
DB 224 TVGFGFNSIMFTIETSLATNLGHLTSKASGVISMAIVG--GALIPPIQGVVTDMLTATE 281  
QY 217 RNALSSMOWSTGLLAAYV--YALSSF 241  
DB 282 SNLLYA--YSVPLLCYFYILFFALKGY 306

Search completed: August 1, 2000, 19:50:26  
Job time: 21506 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:51:34 ; Search time 59.57 seconds  
(without alignments)  
86.206 Million cell updates/sec

Title: US-09-147-955-10  
Perfect score: 1723  
Sequence: 1 MSSSSRRWRENEGRRTLL.....GNLDHEYSLLLEDQDDTPRKP 335

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103.5	6.0	426	5	Patent No. 5268463
2	97	5.6	680	1	US-07-674-287B-2
3	97	5.6	680	2	US-08-436-900A-2
4	96.5	5.6	462	2	US-08-898-976-2
5	96.5	5.6	462	2	US-08-898-976-4
6	95.5	5.5	456	5	5432081-7
7	94	5.5	437	2	US-08-882-704A-6
8	92	5.3	365	1	US-08-390-162-2
9	92	5.3	365	1	US-08-685-945B-2
10	91.5	5.3	415	1	US-08-194-338-10
11	91	5.3	361	1	US-08-390-162-4
12	91	5.3	361	1	US-08-685-945B-4
13	90.5	5.3	214	3	US-08-807-300-3
14	90.5	5.3	238	2	US-08-103-170-13
15	90.5	5.3	303	1	US-08-196-989B-9
16	90.5	5.3	303	2	US-08-760-936-9
17	90.5	5.3	444	1	US-07-781-254A-1
18	90.5	5.3	462	3	US-08-788-231A-15
19	88.5	5.1	444	1	US-07-626-618A-19
20	88.5	5.1	444	1	US-08-333-977-19
21	88.5	5.1	473	1	US-08-597-236-13
22	88.5	5.1	473	1	US-08-746-682A-13
23	87.5	5.1	3033	1	US-07-925-695-5
24	86.5	5.0	443	1	US-07-626-618A-18
25	86.5	5.0	443	1	US-08-333-977-18
26	86	5.0	385	1	US-08-416-756A-3
27	85.5	5.0	348	1	US-08-176-126B-2
28	85.5	5.0	348	2	US-08-669-435-2

```

29 85.5 5.0 348 4 PCT-US94-14431A-2 Sequence 2, Appli
30 85.5 5.0 371 2 US-08-928-692-20 Sequence 20, Appl
31 85.5 5.0 383 1 US-08-196-989B-4 Sequence 4, Appli
32 85.5 5.0 383 2 US-08-760-936-4 Sequence 4, Appli
33 85.5 5.0 682 2 US-08-436-900A-4 Sequence 92, Appli
34 85 4.9 341 2 US-08-846-762-92 Sequence 1, Appli
35 84.5 4.9 433 2 US-08-919-624-1 Sequence 9, Appli
36 84.5 4.9 438 2 US-08-677-049-9 Sequence 20, Appli
37 84 4.9 303 2 US-08-846-762-20 Sequence 92, Appli
38 84 4.9 303 2 US-08-846-762-91 Patent No. 5268463
39 84 4.9 457 5 5268463-7 Sequence 33, Appl
40 83 4.8 184 2 US-08-602-359A-33 Sequence 6, Appli
41 82.5 4.8 388 3 US-08-155-005A-6 Sequence 4, Appli
42 82.5 4.8 407 2 US-08-875-972-4 Patent No. 5432081
43 82.5 4.8 428 5 5432081-9 Sequence 2, Appli
44 82.5 4.8 434 2 US-08-529-600D-2 Sequence 4, Appli
45 82.5 4.8 434 2 US-08-973-275-4

```

## ALIGNMENTS

```

RESULT 1
5268463-8
; Patent No. 5268463
; APPLICANT: JEFFERSON, RICHARD A.
; TITLE OF INVENTION: PLAAANT PROMOTER a-GLUCURONIDASE GENE
; CONSTRUCT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,976
; FILING DATE: 08-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 119,102
; FILING DATE: 10-NOV-1987
; APPLICATION NUMBER: 264,586
; FILING DATE: 31-OCT-1988
; SEQ ID NO:8
; LENGTH: 426
5268463-8

```

Query Match	6.0%	Score 103.5;	DB 5;	Length 426;
Best Local Similarity	21.1%	Pred. No. 0.018;		
Matches 80;	Conservative	63;	Mismatches 133;	Indels 103; Gaps 18;
QY 8	RWRENEGRRTLLGLGLGQLVFDLAIMTFSASLVSTTVDAPLTMSTFTYTWALLYGTI	67		
Db 65	RW-----CKFRPFLFGTAPLMIFD-SVLVFWV-LTOWSHGSKVYVAYLYTMGLGLCYSIV	118		
QY 68	LL-----YRRHKLVPWYVYALLGFDVHGNYLVNKAFTLSITS-VSILDC	113		
Db 119	NIPYGLATAMTQPPQSRARLGAARGTAASLTFFVCL--AFLIGPSIKNSPEEMVSVYHF	176		
QY 114	WTIVMSI-----IFTWMLGKYSYQVFGA-----AICVGGLLLVLS	152		
Db 177	WTIVLATAGWLYFICKSTRENVVRIYAQPSLNISLOTLRNRPFLMCLGA-LCVLIS	235		
QY 153	DSGVTAAGSNPLLDGLFVITGTSILFTLSTVGQE---YC-----VKR--KRIEYVA	198		
Db 236	TFAYSSASSL-----FVVLNDTGLFTVLVQNPWGCGIGTAGAXMVARIGKKTFLGA	289		
QY 199	MIGVFGML-----ISAFETVLERNALSSMOWSTGLLAAY	233		
Db 290	LLGFCGGLLPFWWSLPLVALVALAIASIGQVYTMVWMALEADTVEYGYLTGVRIEG	349		
QY 234	VVYALSSF-----LFCITLPPFLIKMSGAFFNLSDMAVAIVRTFYNDQVDWLY	286		
Db 350	LTYSLSFSTRKCGQAGISIPAFILGLSG----YIANQVQTPVEVINGIKTIALVPCGFM-	405		
QY 287	YLAFLVVGVIYITKTEK 305			
Db 406	LLAFVII--WFFYPLTDK 421			

```

RESULT 2
US-07-674-287B-2
; Sequence 2, Application US/07674287B
; Patent No. 5414076
; GENERAL INFORMATION:
; APPLICANT: Bryan Mark O'Hara
; TITLE OF INVENTION: Gibbon Ape Leukemia
; TITLE OF INVENTION: Virus Receptor
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Karen A. Lowney
; ADDRESSEE: American Cyanamid Company
; STREET: 1937 West Main Street
; STREET: P.O. Box 60
; CITY: Stamford
; STATE: CT
; COUNTRY: USA
; ZIP: 06904-0060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII converted from IBM DW4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/674,287B
; FILING DATE: 19910325
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A., Dr.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 31,104-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203 321 2361
; TELEFAX: 203 321 2971
; TELEX: 710 474 4059
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 680 Amino Acid Residues
; TYPE: AMINO ACID SEQUENCE
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Protein
US-07-674-287B-2

Query Match 5.6%; Score 97; DB 1; Length 680;
Best Local Similarity 21.7%; Pred. No. 0.15;
Matches 66; Conservative 49; Mismatches 105; Indels 84; Gaps 16;

QY 76 LVPWYVALLGFGVDHGNLYNKAFLT-STTSVSLDCW-TIWSIITWMLGTRKYSV 133
|| : || : || : || : || : || : || : || : || : || : || : || : || :
Db 18 LVDYLMMLILGFI-----IAFLAFSGVANDVANSFGTAVGSGVVTLKQACILASI 68
: || : : || : : || : : || : : || : : || : : || : : || : : || :

QY 134 YQFYGAAICVG-----GLLLVLL--SDSGVTAAGS-----NPLL 165
: || : : || : : || : : || : : || : : || : : || : : || : : || :
Db 69 FETGVLLGAKVSETIRKGLIDVEMYNSTQGLLMAGSVSAMFGSAVWQVASFLLKLPIS 128
: || : : || : : || : : || : : || : : || : : || : : || : : || :

QY 166 GDFLVINGSLFTLTSTVGQCYVKRKRIEVV-----AMIGVF-GMLISATEITVLER 217
| : : || : || : || : || : || : || : || : || : || : || : || :
Db 129 GTHCIVGATIGFSLVAKGOE-GVKWSLEIKVMWFVSPLLSGIMSGILFVLRAFILHK 187
: || : || : || : || : || : || : || : || : || : || : || : || :

QY 218 NALSSMOWSGLLAAYVYVYALSSSEFLCT--LTPELLKMSGAAFFNLSMLTSDMNAVALRT 275
|| : || : || : || : || : || : || : || : || : || : || : || :
Db 188 ----ADVPVNGLRALPVFYA-----CTVGINLEFSIMYTGAPLIGFDKL--PLMGITLIS 235
: || : || : || : || : || : || : || : || : || : || : || : || :

QY 276 FIYNQVEDWJLYYLAFLVWVGIFYYTTEK-----DPNNTRALENGNDHEYSLLED 327
: || : || : || : || : || : || : || : || : || : || : || : || :
Db 236 -----VGCAVFCALIVWFVFCPRMKRKIERIKCSPSPESPLMEKKN-----SLKED 281
: || : || : || : || : || : || : || : || : || : || : || : || :

```

D<sub>b</sub> 282 HEET 285

```

RESULT      4
US-08-898-976-2
: Sequence 2, Application US/08898976
: Patent No. 5891670
: GENERAL INFORMATION:
: APPLICANT: Burnham, Martin
: APPLICANT: Lonetto, Michael
: APPLICANT: Warren, Patrick
: TITLE OF INVENTION: NOVEL TETRACYCLINE RESISTANCE
: TITLE OF INVENTION: PROTEIN
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert Price & Rhoads
: STREET: 997 Lenox Drive, Building 3, Suite 210
: CITY: Lawrenceville
: STATE: NJ
: COUNTRY: USA
: ZIP: 08543
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: Fastseq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/898,976
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Bloom, Allen
: REGISTRATION NUMBER: 29,135
: REFERENCE/DOCKET NUMBER: GM10044
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 609-520-3214
: TELEFAX: 609-520-3259
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 462 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-898-976-2

```

Query Match	5.6%	Score	96.5	DB	2	Length	462		
Best Local Similarity	21.5%	Pred. No.	0.098						
Matches	56	Conservative	39	Mismatches	121	Indels	45	Gaps	10
Qy	70	YRRHKFLVPWTWYALLGFVDVHGNYLNVKAKELTSITSVSLD-CWTIVWSIIFTWMFL	127						
Db	17	YSKIKKAVP-----ILLFLFVF-SLVIDNSFKLSVAIADDLNISVTIVSQATLAGLVI	70						
Qy	128	GTKYSVTQFVGAAICV-----GGILLVLLSDSGVTRAGSNPLLGDFLIVNGSILFTLST	181						
Db	71	GMGAVVYASLSDAISRPPTYGVILIFIGSIINFIQHQP-----LTLVGRLIQTAGL	125						
Qy	182	VGOE-----CYCKRKDRIEVAMIG-----VFCMLISATEITVLERNALSSMOWS	226						
Db	126	AAETFLVIYVAKYLSKEDQTYLGLSTSSYSLSIVGTLSGGIFYYL-----HWT	177						
Qy	227	TGLLAAYVYVYALSFLFTLPFLFKMSGAAFFNLSMLTSDMWAVAIFFTYINQVDWLY	286						
Db	178	NMFLIALIVFTLPFLF-KLAPKENNTNKAHLDFGLILVATIATVYMLFITN--FNWLY	234						
Qy	287	YLAFLVVGVIFFYTKTEKDP	307						
Db	235	MIGALIAIVFALYIKNAORP	255						

```

RESULT      5
US-08-898-976-4
; Sequence 4, Application US/08898976
; Patent No. 5891670
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin
; APPLICANT: Lonetto, Michael
; APPLICANT: Warren, Patrick
; TITLE OF INVENTION: NOVEL TETRACYCLINE RESISTANCE
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fastseq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/898,976
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: GM10044
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-898-976-4

```

Query Match	5.6%	Score 96.5;	DB 2:	Length 462;
Best Local Similarity	21.5%;	Pred.	No. 0.098;	
Matches	56;	Conservative	39;	Mismatches 121; Indels 45; Gaps 10;
Qy	70	YRRHKFLVPWVWYAALLGFVDVHGNYLVNKAFELTSITSVSLD-CWTIVWSLIFTVMFL	127	
Dd	17	YSKIKKAVP-----ILLFLFVF-SLVIDNSEKLISVAIADDLNITSVTTVSQOATLAGLVI	70	
Qy	128	GTKYSVVTFQVCAAICV-----GGLLVLSSDSGVTAAGSNPLLGDFLVIPTGSIILETST	181	
Dd	71	GMGAVVTASLDALSRPPFYGVILIFGISIGNIFOHQSP-----UTLVGRIIQTAGL	125	
Qy	182	VQGE-----CYCKRKDRIEVVAMIG-----VFGLMISAETITVLERNALSMMOWS	226	
Dd	126	AAETLVIIYVAKYLKRDKQTYLGLSTSLSIVIGTLSGGFIFTVL-----HWT	177	
Qy	227	TGLLAAYVVYVALSSFECTLPFLFKMGSAFFNLMSLTSDMWAVAIRFTFYNOEVDWL	286	
Dd	178	NMFIALIIVVFTLPFLF-KLLPKENNTNKAHLDFGLLIVATIATTVMLEFITN--FNWL	234	
Qy	287	YLAFCLVVGVIIFYTKTEKDP	307	
Dd	235	MIGALIAIIVPALYIKNAQR	255	

```
RESULT 7
5432081-7
; APPLICANT: JEFFERSON, RICHARD A.
; TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E. COLI
; GLUCONONIDE PERMEASE GENE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,546
; FILING DATE: 15-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 447,976
; FILING DATE: 08-DEC-1989
; APPLICATION NUMBER: 264,586
; FILING DATE: 31-OCT-1988
; APPLICATION NUMBER: 119,102
; FILING DATE: 10-NOV-1987
; SEQ ID NO: 7:
; LENGTH: 456
5432081-7

Query Match 5.5%; Score 95.5; DB 5; Length 456;
Best Local Similarity 20.0%; Pred. No. 0.12;
Matches 76; Conservative 65; Mismatches 136; Indels 103; Gaps 16;

QY 8 RWRENEGMRRTLLGLGOLVSFDLAIMTFSSASLVSTTVDAPLTMSTFTYTVVALLXGTI 67
Db 73 RW-----GKRPPELLFGTAPLMIFSVLVFWV---LTMDSHGSKVYVAYLYTMGLGICYSIV 125
QY 68 LL-----YRRHKFLVPWYVYALLGFDVYHGNVYNKAPELTSITS-VSILDC 113
Db 126 NIPGYSLATMTQPOPSRARGAARGIASLTFVCL--AFLGPKSNSSPEMVSYYHF 183
QY 114 WTIVMSI-----IPTWMLTKYSVYQFVGA-----AICVGGLLLVLLS 152
Db 184 WTIVLAAGMVLVYFCFKSTRENVRIVAQPSLNISLQTLKRNRPFLMCLIGA-LCVLIS 242
QY 153 DSGVTAAGSNPLGDLVLVTGSLFTLSVGOE---YC-----VKKRDIEVV 197
Db 243 TRAVSASS---LFYRVYVNDTGLFTVLVLQVNGVYIGGTAGAMVARIGKKNFTFLG 299
QY 198 AMIGVFQML-----ISATEITVLERNALSSMOWSTGLLAA 232
Db 300 ALLGTCGYLLFFWVWSVSLPVALVALIASIGQGVMTVMWALEADTVGEYLTGVRIE 359
QY 233 YVYVALSSF-----LFTCLTPFLKMSGAAPFNLSMLTSDMMAVAIRTFYINQEDWL 285
Db 360 GLTYSLSFSTRKCGQAGISIPAFILGLSG---YIANQVQTPEVIMGIRTSIALVPCGEM 416
QY 286 YVLAFCVLVVVGFIYTKTEK 305
Db 417 -LLAFVI-----IMFYPLTDK 431

RESULT 7
US-08-82-704A-6
; Sequence 6, Application US/0882704A
; Patent No. 5879906
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Wilson, Katherine J.
; APPLICANT: Leader, Michael
; TITLE OF INVENTION: GLUCONONIDE REPRESSORS AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSES: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,704A
; FILING DATE: 25-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 5879906tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 190106.404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-704A-6

Query Match 5.5%; Score 94; DB 2; Length 457;
Best Local Similarity 18.8%; Pred. No. 0.17;
Matches 82; Conservative 71; Mismatches 125; Indels 158; Gaps 18;

QY 17 RTLLGLGLQ-----LVSFDLAIMTFSSASLVSTTVDAPLTMSTFTYTVV--- 60
Db 8 RTIVGYSGLGVANFAMGALFLLSVYTDVAGVGAAGTML--LLVRVDFADVFAG 65
QY 61 -----ALLYGT--ILLYRRHKFLVPPWV-----YALLGFDVYHGNV-L 95
Db 66 RVVDSVNTRMGKRPPELLFGTAPLMIFSVLVFWVPTDWSHGSKVYVAYLYTMGLGICYSI 125
QY 96 VKAFE--LTSITS-----ISATEITVLERNALSSMOWSTGLLAAVYV 236
Db 126 VNIPGYSLATMTQPOPSRARGAARGIASLTFVCLAFILGPKSNSSPEMVSYYHF 195
QY 115 TIVMSI-----IPTWMLTKYSVYQFVGA-----AICVGGLLLVLLSD 153
Db 186 TIVLAAGMVLVYFCFKSTRENVRIVAQPSLNISLQTLKRNRPFLMCLIGA-LCVLIST 244
QY 154 SGVTAG-----SNPLGDLFVLVTGSLFTLSV---GQCYCVKRDRIEVVAMIG 201
Db 245 FAYSASSLFYRVYVNDTGLFTVLVLQVNLVGTVASAPLVPGMVARIGKKNFTFLGALLG 304
QY 202 VFGML-----ISATEITVLERNALSSMOWSTGLLAAVYV 236
Db 305 TCGYLLFFWVWSVSLPVALVALIASIGQGVMTVMWALEADTVGEYLTGVRIEGLTY 364
QY 237 ALSSF-----LFTCLTPFLKMSGAAPFNLSMLTSDMMAVAIRTFYINQEDWLYLA 289
Db 365 SLFSFTRKCGQAGISIPAFILGLSG---YIANQVQTPEVIMGIRTSIALVPCGEM-LLA 420
QY 290 FCLVVVGFIYTKTEK 305
Db 421 FVI-----IMFYPLTDK 432

RESULT 8
US-08-390-162-2
; Sequence 2, Application US/08390162
; Patent No. 5576192
; GENERAL INFORMATION:
; APPLICANT: Ichikawa, Atsushi
; APPLICANT: Narumiya, Shuh
; TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/390,162  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/024179  
FILING DATE: 23-FEB-1993  
APPLICATION NUMBER: JP 036580-1992  
FILING DATE: 24-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 064889-1992  
FILING DATE: 23-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Fordis, Jean B.  
REGISTRATION NUMBER: 32,984  
REFERENCE/DOCKET NUMBER: 04221-0020-00000  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-390-162-2

Query Match 5.3%; Score 92; DB 1; Length 365;  
Best Local Similarity 19.7%; Pred. No. 0.2;  
Matches 68; Conservative 49; Mismatches 114; Indels 114; Gaps 14;

Qy 4 SSSRRWRENEGRMTLLGLG-----LGQLVSFDLAIMTFSASLVSTVDAP----- 49  
Db 51 SRSYRRRESKRKSKFLCIGLWALTDLVQLTSPVILVLSQRRWEQLDPSGRLCITFF 110  
Qy 50 -LTSFTTYTVVALLYGTLIRRHKFLVPWYVA-----LLGFVDVHGNYLKN 98  
Db 111 GLTM--TVFGSLLLVASAMAVERALAIRAPHNYASHMKTRATPVLLGV-----WLSVL 162  
Qy 99 AFELTSTSVSLDCWTWISITFMFLGT-----KYSVYQFVGAACVGGLL 147  
Db 163 AFALLPVLVGV---RYSVQWP--GTWCFISGPNAGNETDPAREPGSVAFASACILGLLA 217  
Qy 148 LVLLSDSGVTAAGSNPLLDGFLVITGSLTFLSTVGQYCVKR-KDRIEVAMIGVFGML 206  
Db 218 LV-----VTFACNLATIKALVSRCAKAAVSSQSAQWGRI 252  
Qy 207 ISATEITVLERNALSSMOWSTGLLAAYVYVYALSSFLCTLPFLILKMSGAFFNLMS--- 263  
Db 253 TTETAQLMGTCVLSVCWS-----PLIMLMKMFNQMSVEQC 291  
Qy 264 -----LTSMDMAVAIRFTTYNOEVD-WLYYL-----AFCLV 293  
Db 292 KTMGKEKCNSEFLIAVLASLNQILDPPVYLLLRKILLRKFQCI 336

RESULT 9

us-08-685-945B-2  
; Sequence 2, Application 05/08685945B  
; Patent No. 5804415

GENERAL INFORMATION:  
APPLICANT: Ichikawa, Atsushi  
APPLICANT: Narumiya, Shuh  
TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,945B  
FILING DATE: 22-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/024179  
FILING DATE: 23-FEB-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 036580-1992  
FILING DATE: 24-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 064889-1992  
FILING DATE: 23-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Fordis, Jean B.  
REGISTRATION NUMBER: 32,984  
REFERENCE/DOCKET NUMBER: 04221-0020-00000  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-685-945B-2

Query Match 5.3%; Score 92; DB 1; Length 365;  
Best Local Similarity 19.7%; Pred. No. 0.2;  
Matches 68; Conservative 49; Mismatches 114; Indels 114; Gaps 14;

Qy 4 SSSRRWRENEGRMTLLGLG-----LGQLVSFDLAIMTFSASLVSTVDAP----- 49  
Db 51 SRSYRRRESKRKSKFLCIGLWALTDLVQLTSPVILVLSQRRWEQLDPSGRLCITFF 110  
Qy 50 -LTSFTTYTVVALLYGTLIRRHKFLVPWYVA-----LLGFVDVHGNYLKN 98  
Db 111 GLTM--TVFGSLLLVASAMAVERALAIRAPHNYASHMKTRATPVLLGV-----WLSVL 162  
Qy 99 AFELTSTSVSLDCWTWISITFMFLGT-----KYSVYQFVGAACVGGLL 147  
Db 163 AFALLPVLVGV---RYSVQWP--GTWCFISGPNAGNETDPAREPGSVAFASACILGLLA 217  
Qy 148 LVLLSDSGVTAAGSNPLLDGFLVITGSLTFLSTVGQYCVKR-KDRIEVAMIGVFGML 206  
Db 218 LV-----VTFACNLATIKALVSRCAKAAVSSQSAQWGRI 252  
Qy 207 ISATEITVLERNALSSMOWSTGLLAAYVYVYALSSFLCTLPFLILKMSGAFFNLMS--- 263  
Db 253 TTETAQLMGTCVLSVCWS-----PLIMLMKMFNQMSVEQC 291  
Qy 264 -----LTSMDMAVAIRFTTYNOEVD-WLYYL-----AFCLV 293

Db 292 KTMGKEKCSFLIAVRLASNLQILDPPWYLLLRKILLRKFCQI 336

## RESULT 10

US-08-194-338-10

; Sequence 10, Application US/08194338

; Patent No. 5474898

; GENERAL INFORMATION:

; APPLICANT: Venter, John C.

; APPLICANT: Fraser, Claire M.

; TITLE OF INVENTION: OCTOPAMINE RECEPTOR

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe, Martens, Olson and Bear

; STREET: 620 Newport Center Drive, Sixteenth Floor

; CITY: Newport Beach

; STATE: CA

; COUNTRY: USA

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/194,338

; FILING DATE: 08-FEB-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/676,174

; FILING DATE: 28-MAR-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Israelsen, Ned A.

; REGISTRATION NUMBER: 29,655

; REFERENCE/DOCKET NUMBER: NIH101.001DV1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 415 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: Internal

US-08-194-338-10

Query Match 5.3%; Score 91.5; DB 1; Length 415;  
Best Local Similarity 19.5%; Pred. No. 0.26;  
Matches 64; Conservative 36; Mismatches 95; Indels 133; Gaps 14;

QY 78 PWYVY-----ALGFVDVHGNYLV-----NKAFELTS---ITSVSILDC-----WT 115

Db 32 PHNYVAMLLTLDFIVFGNVLCMAVSREKALQTTTYLVLSLAVADLLVATLVMPV 91

QY 116 IVMSIITFMFLGTRKYSVYQFGAAICVGGLLLVLLSDSGVTAAGSNPLLGDFLVITGSI 175

Db 92 VLEVVGKFKSRHIDIF-----VTIL-----DVMCMCTASI 122

QY 176 LFTLSTVGQYCVKRDREIVAMIGVFGMLISATBITVLERNALSSMQWSTGLLAAYV 235

Db 123 L-----NLCAISIDRYTAVAM-----PMLNTRYSSKRRVTVMIA--IV 159

QY 236 YALSSFLCTLTPLFLKMSGAPFNLSMLTSDMWAIVARTFIYNQEVDMLY--YLAFCLV 293

Db 160 WLSFTSITCPL-----IFQLNNTDQNECIIANFAVYVSSIVSFVFFITVLV 208

QY 294 VVGIFYITKTEKDPNTR-----ALENGNLD----- 319

Db 209 YIKIYIVLRKRKRVTNKKSSRAFRANLKTPLKDAARRAGEMEMLSTTAPPERTRYS 268

QY 320 -----HEYSL-----LEDQDDTPRP 335

Db 269 IPPSHQLTLPDPSSHGHLSNPDSPAKP 296

## RESULT 11

US-08-390-162-4

; Sequence 4, Application US/08390162

; Patent No. 5576192

; GENERAL INFORMATION:

; APPLICANT: Ichikawa, Atsushi

; APPLICANT: Narumiya, Shuh

; TITLE OF INVENTION: Prostoglandin E Receptors, Their DNA and

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; ADDRESSEE: Dunner

; STREET: 1300 I Street, N.W., Suite 700

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/390,162

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/024179

; FILING DATE: 23-FEB-1993

; APPLICATION NUMBER: JP 036580-1992

; FILING DATE: 24-FEB-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 064889-1992

; FILING DATE: 23-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Fordis, Jean B.

; REGISTRATION NUMBER: 32,984

; REFERENCE/DOCKET NUMBER: 04221-0020-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-408-4000

; TELEFAX: 202-408-4400

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 361 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-390-162-4

Query Match 5.3%; Score 91; DB 1; Length 361;  
Best Local Similarity 19.8%; Pred. No. 0.25;  
Matches 68; Conservative 48; Mismatches 113; Indels 114; Gaps 14;

QY 4 SSSRRRENEGRRTLLIGL-----LQLVSEFDIAIMTFFSASLIVSTTVDP----- 49

Db 51 SRSYRREKRKRKSLFCIGWLALTDLVQLLTSPVILVLSQRWEQLDPSGRICTF 110

QY 50 -LTMSTFTTVVVALYGTILYRRHFLVPWYVA-----LLGFVDVHGNYLVNK 98

Db 111 GLTM--TVFGLSLLVASMAVERALAIRAPHWASHMKTRATPVLGV-----WLSVL 162

QY 99 AFELTSITSVIDCWTIVMSIIFTWMFLGT-----KYSVYQFYGAICVGGLL 147

Db 163 AFALLPVLGVG---RYSVQWP--GTWCFFSTGPAGNETDPAREPGSVAFASAFACLAGLLA 217  
QY 148 LVLLSDSGVTAAGSNPLLGDFLVTGSLFTLSTVGVQECVKR-KDRIEVVAMIGVFGML 206  
Db 218 LV-----VTFACNLATIKALVSRCAKAAVSSQAQNGRI 252  
QY 207 ISATEITVLERNALSSMOWSTGLAAVYVVALSSFLCTLTPLFLKMSGAAFFNLSM--- 263  
Db 253 TTETAIQLMGIMCVLSVCWS-----PLIMLMKMFNQMSVEQC 291  
QY 264 -----LTSMDMAVAIRTFIYNOEVD-WLYYL-----AFC 291  
Db 292 KTMGKEKCNCSFLIAVRLASLNQILDPPWYLLLRKILLRRKFC 334  
RESULT 12  
US-08-685-945B-4  
; Sequence 4, Application US/08685945B  
; Patent No. 5804415  
; GENERAL INFORMATION:  
; APPLICANT: Ichikawa, Atsushi  
; TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 22-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/024179  
; FILING DATE: 23-FEB-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 036580-1992  
; FILING DATE: 24-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 064889-1992  
; FILING DATE: 23-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fordis, Jean B. 32,984  
; REGISTRATION NUMBER: 32,984  
; REFERENCE/DOCKET NUMBER: 04221-0020-02000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 361 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-685-945B-4

Query Match 5.3%; Score 91; DB 1; Length 361;  
Best Local Similarity 19.8%; Pred. No. 0.25; Indels 114; Gaps 14;  
Matches 68; Conservative 48; Mismatches 113;

QY 4 SSSRRRENCGMRRLLGLG-----LGQVLSFDLAIMTFASLSVSTVDPAD----- 49  
Db 51 SRSYRRRESKRKKSFLLCIGHWAULTDLVGQLLTSPPVILVYLSQRRWEQLDPSGLCTFF 110

QY 50 -LTMSTYTTVVALYGTILLYRRHKFLVPWYTA-----LLGFDVHGHVLYNK 98  
Db 111 GLTM--TVFGSSLLVASAMAVRALAIRAPHWTASHMKTRATPVLLGV-----WLSVL 162  
QY 99 AFELTSITSVILDCWTVMSIIFTWMLGT-----KYSVYQFVGAAICVGGLL 147  
Db 163 AFALLPVLGVG---RYSVQWP--GTWCFFSTGPAGNETDPAREPGSVAFASAFACLAGLLA 217  
QY 148 LVLLSDSGVTAAGSNPLLGDFLVTGSLFTLSTVGVQECVKR-KDRIEVVAMIGVFGML 206  
Db 218 LV-----VTFACNLATIKALVSRCAKAAVSSQAQNGRI 252  
QY 207 ISATEITVLERNALSSMOWSTGLAAVYVVALSSFLCTLTPLFLKMSGAAFFNLSM--- 263  
Db 253 TTETAIQLMGIMCVLSVCWS-----PLIMLMKMFNQMSVEQC 291  
QY 264 -----LTSMDMAVAIRTFIYNOEVD-WLYYL-----AFC 291  
Db 292 KTMGKEKCNCSFLIAVRLASLNQILDPPWYLLLRKILLRRKFC 334  
RESULT 13  
US-08-807-300-3  
; Sequence 3, Application US/08807300  
; Patent No. 6010852  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: NOVEL HUMAN PROTON ATPASE SUBUNIT  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/807,300  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0222 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 214 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1132516  
US-08-807-300-3

Query Match 5.3%; Score 90.5; DB 3; Length 214;  
Best Local Similarity 21.5%; Pred. No. 0.14;  
Matches 47; Conservative 32; Mismatches 59; Indels 81; Gaps 11;





```
QY 176 LFTLSTVGQCYCVKRRKDRIEYVAMIGVFGMLISATEITVLERNALSSMOWSTGLAAVYV 235
Db 123 L-----NLCAISIDRYTAVAM-----PMLNTRYSSKRRVTVMIA--IV 159
QY 236 YALSSFLFCTLTPELLKMSGAAFFNLSTDMWAVAIRTFIYNQOVDWLY--YLAFCV 293
Db 160 WLSFTISCPL-----LFLNNTDQNECIIANPAFVYSSIVSFYVPFIVTLV 208
QY 294 VVGIFITYKTEKDPNTR 311
Db 209 YIKIYIVLRKRKRNTK 226
```

Search completed: August 1, 2000, 19:51:36  
Job time: 18758 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: August 1, 2000, 21:31:37 ; Search time 61.65 seconds  
(without alignments)  
376.755 Million cell updates/sec

Title: US-09-147-955-10  
Perfect score: 1723  
Sequence: 1 MSSSSRRWRENEGMRRTLL.....GNLDHEYSLEDDQDTPRKP 335

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

SPTREMBL\_12:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	428	24.8	505	3	O59785 schizosacch
2	175	10.2	4	Q14963	Q14963 homo sapien
3	134	5	292	1	O59388 pyrococcus
4	125	7.3	370	3	Q04083 saccharomyc
5	122	7.1	341	3	O94654 schizosacch
6	120	7.0	411	10	P93390 nicotiana t
7	117	6.8	408	10	P92991 arabidopsis
8	114	6.6	287	2	O06933 malonomonas
9	114	6.6	311	1	Q9Y8U7 aeropyrum p
10	112	6.5	294	2	Q9WFO0 thermotoga
11	111	6.4	225	5	O96702 ilyanassa o
12	110	6.4	340	10	O82600 arabidopsis
13	110	6.4	382	4	Q9Y308 homo sapien
14	110	6.4	445	2	O34383 bacillus su
15	109	6.3	443	8	Q34844 katharina t
16	107	6.2	394	10	O24091 medicago tr
17	106	6.2	390	10	P93642 zea mays
18	105	6.1	378	2	O67513 aquifex aeo
19	105	6.1	434	3	O94695 schizosacch

20	105	6.1	548	8	O63543 bodo saltan
21	105	6.1	617	1	O59120 pyrococcus
22	104	6.1	479	2	O87134 vibrio chol
23	104	6.1	479	2	O87155 vibrio chol
24	104	6.0	275	1	O57793 pyrococcus
25	104	6.0	362	10	O82750 arabidopsis
26	103	6.0	344	2	O66073 bacillus li
27	102	5.9	285	1	O58347 pyrococcus
28	102	5.9	301	5	O19026 caenorhabdi
29	102	5.9	410	10	P93389 nicotiana t
30	102	5.9	485	2	Q926M8 chlamydia p
31	102	5.9	296	1	Q9Y625 aeropyrum p
32	101	5.9	273	2	Q9X2A7 thermotoga
33	101	5.9	1001	11	O62439 mus musculu
34	101	5.9	1028	6	O18886 oryctolagus
35	100	5.8	328	10	O81326 arabidopsis
36	100	5.8	459	8	O21710 pygathrix r
37	100	5.8	755	8	Q9XMS0 tetrahymena
38	99	5.8	364	5	O20787 caenorhabdi
39	99	5.8	457	8	O47876 alligator m
40	99	5.8	935	3	O42986 schizosacch
41	99	5.7	284	1	O28298 archaeoglob
42	99	5.7	291	2	O86204 escherichia
43	99	5.7	406	4	Q9Y522 homo sapien
44	99	5.7	745	2	Q924C1 escherichia
45	99	5.7	745	2	Q9WW95 escherichia

## ALIGNMENTS

RESULT 1  
O59785 PRELIMINARY: PRT; 505 AA.  
ID O59785  
AC O59785;  
DT 01-AUG-1998 (TREMREL. 07, Created)  
DT 01-AUG-1998 (TREMREL. 07, Last sequence update)  
DE 01-AUG-1998 (TREMREL. 07, Last annotation update)  
DE HYPOTHEICAL 56.8 KD PROTEIN.  
GN SPCC320.08.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972H-;  
RA WOOD V., RAJANDREAM M.A., BARRELL B.G., WEDLER H., WAMUTT R;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL022245; CAA18310.1;  
KW Hypothetical protein.  
SQ SEQUENCE 505 AA; 56780 MW; 71E0677A CRC32;

Query Match	24.8%;	Score 428;	DB 3;	Length 505;
Best Local Similarity	31.4%;	Pred. No. 6.6e-22;		
Matches	93;	Conservative	65;	Mismatches 108; Indels 30; Gaps 7;
QY	24	LGQVSEDLAIMTFSSASLVSTVDAPLMTSFTTYTVVALLYGTLTYR-----R	72	
Db	147	LGQVLSLCITATNFNGMSGISNIPAFQFLVYALLTLVYTPYVFMGFKYFEMIFR	206	
QY	73	HKFLVPWY-YALLGFVDVGNVLYNKAFLTSITSVSLDCWTIVMSIITWMLTKY	131	
Db	207	HG-----WKYIIFAFDFVEGNYFVFLAYQYTNMLSASLDSNATVAVVILSF	260	
QY	132	SVYQFVCAATCVGGLLVLLSDSGVTAAGS---NPLIGDFLVITGSLTFLSTVG	187	
Db	261	HWSQILGVVACIGLVLLVVSVD--VISRGDYSAVNPGDGGVITIGATCYG	318	
QY	188	VKRDRTEVAMIGVFGMLISATEITVLERNALSSMOWST---GLLAAYVYVAL	244	
Db	319	ASKPLVAVIGQLSGLYSIISITQTFIDRHLTYLHWTSEMGYLAGFI---LVN	375	





Db 97 DMSGKLTVEGLGLFAMVYLFNIYNIKNQVLK-ALHAPMTVTLVQPAVGSVLI-TIM 154  
QY 69 ----LYRRHKE-----LVPWYVALLGVVDVHGNYLVNKAFELTSITSVSLDCWTIV 117  
Db 155 WYUNLTKRKPRISGAQAALILP-----LAVVHTLGLNFTNMSGKVSFSFTHIKAMEPF 208  
QY 118 WSIITWFMFLGTYKYVYQVGAACVCGLLLVLLSDSGVTAAGSNPLLDGLDFLVITGSILF 177  
Db 209 FSVLLSAMFLGKPTWVLGAIVPIVGGVALASISEVSNWAGFSAMASNLTNQ----- 263  
QY 178 TLTSTVQGYCYCKRKRIEVAMIGVFMGLISATEITVTLERNALSSMOWSTGLLAAYVYVA 237  
Db 264 SRNVLSKKVMVKDDSLDNLITFSI-----ITLMSL-----VLMADPVTFF 303  
QY 238 LSSFLFCTLTPLFLKMSGAFFNLSMLTSDMVAVALRTFYNOEVDWL-----YYL 288  
Db 304 TEGIKF---TPSYIQAGV---NVKQIYTKSLIALCFAHAY-OOVSYMILARVSPVTHSV 356  
QY 289 AFC-----LVVVGIFITKTEKDPNN 309  
Db 357 GNCVKRVVIVSSVIEFKTPVSPVN 381  
RESULT 8  
ID O06933 PRELIMINARY; PRT; 287 AA.  
AC O06933;  
DT 01-JUL-1997 (Tremblrel. 04, Created)  
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
DE MADN.  
GN MADN.  
OS Malonomonas rubra.  
OC Bacteria; Malonomonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97274645.  
RA BERG M., HILBI H., DIMROTH P.;  
RT "Sequence of a gene cluster from Malonomonas rubra encoding components  
RT of the malonate decarboxylase Na+ pump and evidence for their  
RT function.";  
RL Eur. J. Biochem. 245:103-115(1997).  
DR EMBL; U87980; AAC45409.1; -.  
SQ SEQUENCE 287 AA; 31791 MW; 6B3209C0 CRC32;

Query Match 6.6%; Score 114; DB 2; Length 287;  
Best Local Similarity 21.5%; Pred. No. 0.73;  
Matches 65; Conservative 65; Mismatches 122; Indels 50; Gaps 14;

QY 32 LAITMSTTVDAPLTMSFTTYVVALLYGTYLLYRRHKFLVPMWYVALLGF 87  
Db 4 LTMTFVAFSFLIGYLAGVDAYFSVLIRVSLAALFPLPMRSVSKIALKMLVIGG 63  
QY 88 VDVGNYL-VNKAFLTSITSVSLDCWTIVSIITWTF--LGTKSYVYQVGAACVIG 144  
Db 64 IQGCMVYVYQSFLLTVPEVLI---FILTPYVTLFIDLMQGCFSRHYLLTAVLV 120  
QY 145 GLLVLLSDSGVTAAGSNPLLDGLFVITGSILFSLTSTVQGYCYCKRKRIEVAMIGVFG 204  
Db 121 GAFIIRYSN-----VGDNVLAGFVVGQANICFALQGVYKLLIER-EKLE-LSQRSVFG 173  
QY 205 -----MLISATEITVL-----ERNALSSMOWSTGLLAAYVYVVALSSFLFCTLTPLFLKMSG 255  
Db 174 YFVLGALIIATAWAIFGNPEKLPATSLQGLVLYIGVLTASGIGYFLW-----NKG 224  
QY 256 A-----AFFNLSMLTSDMVAVALRTFYNOEVDWL-YLAFCLVYVVGIFYT---KT 303  
Db 225 ATKVDGVGLAINNALIPA---GLIVNLLIWNEDADLVRLFTGGSIALLSLIHFWGWT 281  
QY 304 EK 305  
RP SEQUENCE FROM N.A.

Db 282 KK 283  
RESULT 9  
ID Q9Y807 PRELIMINARY; PRT; 311 AA.  
AC Q9Y807;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE 311AA LONG HYPOTHETICAL PROTEIN.  
GN APE2537.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Aeropyrum.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K1;  
RX MEDLINE; 99310339.  
RA KAWABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,  
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,  
RA HOSUYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,  
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,  
RA YAMAZAKI J., KOSHIDA N., OGUCHI H., AOKI K., KUBOTA K., NAKAMURA Y.,  
RA NOMURA N., SAKO Y., KIKUCHI H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
DR EMBL; AP000064; BA081553.1; -. 15345528 CRC32;  
SQ SEQUENCE 311 AA; 32703 MW; 15345528 CRC32;

Query Match 6.6%; Score 114; DB 1; Length 311;  
Best Local Similarity 20.3%; Pred. No. 0.79;  
Matches 62; Conservative 51; Mismatches 127; Indels 66; Gaps 11;

QY 32 LAITMSTTVDAPLTMSFTTYVVALLYGTYLLYRRHKFLVPMWYV 82  
Db 9 LALAVANVSTASILVRLAGVHGFAAATWRLVLTILTAALVAAVAKKHKRPVLP---- 64  
QY 83 ALLGFVD-----VHGNYLVNKAFLTSITSVSLDCWTIVSIITWFMFLGTK 130  
Db 65 ---GGVLLLVLAASGVVALAHFDLNMASLFHMSVAMSVTVVDSYPAVLAIVGRFLF-KEK 120  
QY 131 YSVYQVFGAAICVGGLL-LVLLSDSGVTA-AGSNPLLDGLFVITGSILFSLTSTVGOEYCV 188  
Db 121 YTPQLLGLAVIAMVGVAGTAFYSRCSAPPDGPDKGFLAFAGML-----AVAAVFSI 175  
QY 189 -----KKDRIEVYAMIGVFMGLISATEITVTLERNALSSMOWSTGLLAAYVYVVALSSFLF 243  
Db 176 GKGLRSKYSTLEVTLPVYGFAAIVSLTLTLVVVRAPL-----TG-----YTAETLY 222  
QY 244 CILTPELLKMSGAAFFNL-----SMLTSDMVAVALRTFYNOEVDWL-YLAFCL 291  
Db 223 LILALLPLMGLGHTLINLLRLRLSLAATVPILGEPVGAALAWILLGEKVTLEALLMV 282  
QY 292 LVVVG 297  
Db 283 VVLGI 288

RESULT 10  
ID Q9WYQ0 PRELIMINARY; PRT; 294 AA.  
AC Q9WYQ0;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE SUGAR ABC TRANSPORTER, PERMEASE PROTEIN.  
GN TM0419  
OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga.  
RN [1]  
RP SEQUENCE FROM N.A.



```

QY 143 VGGLLVL-----LSDSGVTAAGSNP-----LLGDFLVITGSIILF-TLSTVGQCYCVK 189
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 124 VGLLLTWTGQPLTNSPEQAAANSNNHCHENWIKGCFLLGTGVLFSSWMLIAQKINKV 183
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 190 RKDRIEWAMIGVFGML-----ISATEITVLERNALSMQWSTGLLAAY 233
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 184 YPCPYSTVILSVFGTGLQCALLSLIKTRHLEDWILRDELTI-----TVVIAGV 232
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 234 VVALSSFLFCLT-PPFLKMSG---AAFNLSMLTSDMWAVAIRTFIYNQVDMWLYLA 289
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 233 VAQGM-----CTVGMSCWICKQGPVVSFSFSPVVLMS---ATVDFDLHLHREIYLSGVIC 284
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 290 FCLVVVGIFV-----TKTEKDPNNT 310
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 285 SVVVIGLYIFLWSRSQKIVECKIMKLPNT 315
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 13
QY308
ID QY308 PRELIMINARY; PRT; 382 AA.
AC QY308
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE CGI-19 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RA LIN W.-C.;
RT "Comparative gene cloning: Identification of novel human genes with C.
RT elegans proteome as template.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF139553; AAD27728.1; -.
SQ SEQUENCE 382 AA; 42458 MW; 65BFEC68 CRC32;

Query Match 6.4%; Score 110.5; DB 4; Length 382;
Best Local Similarity 21.3%; Pred. No. 1.7; Mismatches 102; Indels 67; Gaps 13;
Matches 60; Conservative 53;

QY 52 MSPTTYVVVALLYGTYLLRRHKFLVPWYALLGPDVGHVGNLVNKAPELTSITSVIL 111
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 118 VQFAFYSIFGLIELQIQDKRRR--IPGKTYMIIAELTGTWGLSNTSLGYNLPQVIF 175
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 112 DCWTIVWSIIFTMFLGTYSVYQVGAICVG-GLLLVLLSDSGVTAAGSNPLLGDFLV 170
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 176 KCCKLIPVMLGGVFIOGKRYNAD-VSAATCMISGLIWFETLADS-TTAPNFN----- 225
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 171 ITGSILFTLSTVGQECVVRKORIEVVMGVGMLISATEITVLERNALSS--MQWSTG 228
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 226 LTGVLLISALCAD-----AVIG-----NVOEKAMKLNHNASDMLVYSYS 266
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 229 LLAAYVVVYALS-----SFLFCTLP-----FLKMSGAAFFNLSMLTS-----DM 268
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 267 IGFVYILLGLTCTSGLGPATVCAKNPVTYGVAFLSLTG--YFGISFVLALIKIFGAL 324
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 269 WAVAIRF-----FTYNQVDMWLYLAFCLVVGIFV 299
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 325 IAVTVTGRKAMTIVLSIFFFAKPFQYVWVMSGLLVVLGIFL 366
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 14
Q34383
ID Q34383 PRELIMINARY; PRT; 445 AA.
AC Q34383;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)
DE PUTATIVE TRANSPORTER.
GN YOCR.

```

```

OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RA LAPIDUS A., GALLERON N., SOROKIN A., EHRLICH D.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE; 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA FRITZ C., FUJITA M., FABRET C., FERRARI E., FOULGER D.,
RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISEPPI G., GUY B.J., HAGA K., HALECH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLSAPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA JORIS B., KARAWATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA NOONE D., O'REILLY M., OGAWA K., OGWARA A., OUDEGA B., PARK S.H.,
RA PARKO V., POHL T.M., PORTELLE D., PORWOLLIK S., PRESCOTT A.M.,
RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SEKIGUCHI J., SEKOWSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDI B.,
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEWAKU K.,
RA TAKUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMBUITT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF027868; AAB84480.1; -.
DR EMBL; 299114; CAB13826.1; -.
DR PFAM; PF00209; SNF; 2.
SQ SEQUENCE 445 AA; 48334 MW; 1E7E576A CRC32;

```

```

Query Match 6.4%; Score 110.5; DB 2; Length 445;
Best Local Similarity 20.6%; Pred. No. 2;
Matches 62; Conservative 48; Mismatches 104; Indels 87; Gaps 14;

QY 6 SRWRNEGRRRTLLG--LGLGOLVSFDLAIMTFSASLSTVTDAPLTSFTTYTV---V 60
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 4 ANQWSSKGLFILAAGSAIGLGAIRKFPY-----VAGTSGGGAFLIFLFTVLIGL 55
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 61 ALLYGILLVRR-----HKPLVP---WYVALLGFDVGHVGNLVNKAPELTSITSVS 109
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 56 PLLLGEFIIGRTQKDAVQSYKTLPAGTKHWHIGYLGWV----- 94
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 110 ILDCWTIV--WSIIFTMFLGTYSVYQVGAICVGGALLLVLLSDSGVTAAGSNPLGD 167
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 95 --TCFILLSFYSVGGWILI---YIIKGTGGLSQTSGF-----DTIFASTISNP--- 139
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 168 FLVITGSILFTLSTV-----GOEYCVKRDRIEVVAMIGVFGMLI--SATEITVLERNAL 220
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 140 YLAVGQLLFVWVTVILVAKGVSSGIEKASQVLMFALFLFILWVRSVTLDGAMEGKIF 199
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

```



Search completed: August 1, 2000, 21:31:42  
Job time: 6071 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:32:20 ; Search time 31.57 Seconds  
(without alignments)  
328.836 Million cell updates/sec

Title: US-09-147-955-10  
Perfect score: 1723  
Sequence: 1 MSSSSRRWRENGMRRTLL.....GNLDHEYSLEDDQDDTPRKP 335

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues  
Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_38:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	145	8.4	324	1 YAGC_SCHPO	Q9875 schizosacch
2	128	7.4	304	1 PAGO_SALTY	O30646 salmonella
3	124.5	7.2	305	1 YBVB_BAGSU	O32256 bacillus su
4	124	7.2	330	1 YETK_BAGSU	O31540 bacillus su
5	121	7.0	306	1 YEDA_ECOLI	P09185 escherichia
6	119	6.9	490	1 CAR3_DICDI	P35352 dictyosteli
7	109.5	6.4	541	1 YD18_MYTU	Q10631 mycobacteri
8	108.5	6.3	330	1 P355_SYNY3	P74436 synechocyst
9	106.5	6.2	531	1 PTM1_YEAST	P32857 saccharomyc
10	106	6.2	402	1 CPT2_BRAOL	P52178 brassica ol
11	105	6.1	305	1 YBFH_BAGSU	O31448 bacillus su
12	105	6.1	549	1 COX1_TRYBB	P04371 trypanosoma
13	105	6.1	1021	1 TSCC_HUMAN	P35017 homo sapien
14	103.5	6.0	311	1 YCAY_CLOKL	P38943 clostridium
15	103.5	6.0	393	1 YMB8_YEAST	Q38730 saccharomyc
16	103.5	6.0	637	1 YHE7_YEAST	P38724 saccharomyc
17	102.5	5.9	686	1 NU5M_SCHCO	P50368 schizophyll
18	102	5.9	411	1 PE23_RABIT	P46069 oryctolagus
19	102	5.9	517	1 QOX1_SULAC	P38004 sulfolobus
20	101.5	5.9	312	1 OLF2_CHICK	P37068 gallus gall
21	101	5.9	459	1 NU4M_PELSU	O79677 pelomedusa
22	100.5	5.8	318	1 OLF1_CHICK	P37067 gallus gall
23	99.5	5.8	461	1 PUCC_RHOCA	P23462 rhodobacter
24	99.5	5.8	970	1 Y277_MYCGE	Q49409 mycoplasma
25	99	5.7	432	1 Y672_METJA	Q38086 methanococc
26	99	5.7	474	1 NU4M_DIDMA	P41308 didelphis m
27	99	5.7	1105	1 YEGE_ECOLI	P38097 escherichia
28	98.5	5.7	339	1 NUOH_RICPR	Q98cf7 rickettsia
29	98.5	5.7	444	1 CADB_ECOLI	P38891 escherichia
30	98.5	5.7	488	1 LYSP_ECOLI	P25737 escherichia
31	98	5.7	301	1 YIJE_ECOLI	P32667 escherichia
32	97.5	5.7	407	1 GLUP_HELPY	O25788 helicobacte
33	97.5	5.7	409	1 NU4M_CAEEL	P24892 caenorhabdi

34	97.5	5.7	572	1 NU5M_LOOMI	Q36428 locusta mig
35	97	5.6	273	1 BACA_ECOLI	P31054 escherichia
36	97	5.6	343	1 Y841_METHA	O26929 methanobact
37	96	5.6	326	1 UGNT_CANFA	O77592 canis famil
38	96	5.6	574	1 NU5M_DROME	P18932 drosophila
39	95.5	5.5	340	1 COX1_CRION	P98003 crithidia o
40	95.5	5.5	444	1 D2DR_BOVIN	P20288 bos taurus
41	95.5	5.5	448	1 PCAK_PSEPU	O51955 pseudomonas
42	95.5	5.5	609	1 NAH_CAEEL	P35449 caenorhabdi
43	95.5	5.5	893	1 PMTX_SCHPO	O13898 schizosacch
44	95	5.5	289	1 Y510_ARCFU	O39740 archaеoglob
45	95	5.5	580	1 NU5M_ANOGA	P34854 anopheles g

## ALIGNMENTS

RESULT 1	
YAGC_SCHPO	
ID YAGC_SCHPO	STANDARD; PRT; 324 AA.
AC Q09875;	1996 (Rel. 33, Created)
DT 01-FEB-1996	(Rel. 33, Last sequence update)
DT 01-FEB-1996	(Rel. 33, Last annotation update)
DE HYPOTHETICAL 35.8 KDA PROTEIN C12G12.12 IN CHROMOSOME I.	
GN SPAC12G12.12.	
OS Schizosaccharomyces pombe (Fission yeast).	
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;	
OC Schizosaccharomycetaceae; Schizosaccharomycetes.	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=972;	
RA Devlin K., Odell C., Churcher C.M., Barrell B.G., Rajandream M.A.,	
RA Walsh S.V.;	
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.	
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).	
CC -----	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC or send an email to license@isb-sib.ch).	
CC -----	
DR EMBL; Z66568; CAA91507.1; --	
FT Hypothetical protein; Transmembrane.	
FT TRANSMEM 34 54	POTENTIAL.
FT TRANSMEM 76 96	POTENTIAL.
FT TRANSMEM 103 123	POTENTIAL.
FT TRANSMEM 127 147	POTENTIAL.
FT TRANSMEM 158 178	POTENTIAL.
FT TRANSMEM 198 218	POTENTIAL.
FT TRANSMEM 243 263	POTENTIAL.
FT TRANSMEM 275 295	POTENTIAL.
FT TRANSMEM 297 317	POTENTIAL.
SQ SEQUENCE 324 AA; 35746 MW; 9696681A0E82CCD3 CRC64;	

Query Match 8.4%; Score 145; DB 1; Length 324;  
Best Local Similarity 23.8%; Pred. No. 0.00087;  
Matches 67; Conservative 42; Mismatches 117; Indels 56; Gaps 10;

Qy	68	LLYRRHKFLVPWY-----	WYALLGFVDVGHVYLVNKAFELYSIVSVILDCWT 115
Db	53	VLYKRHS-QGPGYESLDHLPLKHKVFMAIPALMDICGSLTMNVGLLYTSASIYQTRGSL 111	
Qy	116	IVWSIIFTWFMELGTYKYVQFVGAICVGLLLVLLSDSGVTAAGSNPLIGDFLVTGSI 175	
Db	112	IFVAFVATTLTKRTIGQLQMLSFVVLGVAIVGYSGSS-SSIGSNPLIGTAVLIGQF 170	
Qy	176	LFTLSTVGQYCVK--RKDRIEVAVAMIGVFGMLISATITVLERNALSSMOWSTGLAAY 233	

Db 171 FLATQFTIEEYILSIQVDPSELVAYEGYGVFVLGMII-----SYFIGSTTAGY 223  
 QY 234 -----VYALSSFLCTLTFLKMSGA-----AFNLMSL-----T 265  
 Db 224 HGFEDSHVISRF---NEVPALVIVSGVILVSTAFNTVSGLAITKLHSATRSLLDIART 280  
 QY 266 SDMWAVAIRTFIYNQEDWLYLAFCLVGVVGFYIYTKTEKDP 307  
 Db 281 FGIWIAMANGM--ESFHLQFLGFLVLLIYGIYTHSIKFP 320

RESULT 2  
 PAGO\_SALTY STANDARD; PRT; 304 AA.  
 AC PAGO\_SALTY 30646;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE PAGO PROTEIN.  
 GN PAGO.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 140285;  
 RA Gunn J.S., Belden W.J., Miller S.I.;  
 RT Identification of Phop-PhoO activated genes within a duplicated  
 RT region of the Salmonella typhimurium chromosome.";  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC 1- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF013775; AAB82452.1; .  
 CC DR STYGENE; SG77777; PAGO.  
 CC DR PFAM; PF00892; DUF6; 2.  
 KW Transmembrane.  
 FT TRANSMEM 4 24 POTENTIAL.  
 FT TRANSMEM 34 54 POTENTIAL.  
 FT TRANSMEM 67 87 POTENTIAL.  
 FT TRANSMEM 95 115 POTENTIAL.  
 FT TRANSMEM 119 139 POTENTIAL.  
 FT TRANSMEM 150 170 POTENTIAL.  
 FT TRANSMEM 180 200 POTENTIAL.  
 FT TRANSMEM 214 234 POTENTIAL.  
 FT TRANSMEM 246 266 POTENTIAL.  
 FT TRANSMEM 267 287 POTENTIAL.  
 SQ SEQUENCE 304 AA; 33622 MW; B368153FD6EBFB0B CRC64;

Query Match 7.4%; Score 128; DB 1; Length 304;  
 Best Local Similarity 21.9%; Pred. No. 0.013;  
 Matches 61; Conservative 53; Mismatches 119; Indels 46; Gaps 10;

QY 59 VVALLYGTILLY----RRHKFLVPWYVYALLGFDVGHVGNLVNKAFLTSITSVSLDCW 114  
 Db 50 IIAWLKKTLFFPGQRLQFVICIYFCIPFSLMIYGETYVNSGLAAIIFANMPVA--- 106  
 QY 115 TIWVSIFTMFLGTIKYSYQFVGAACVGLLLVLLSDS-----GVTAAGSNPLL 165  
 Db 107 -----VLIASVFLNEKAKLQIAGLTAITALTGILLEENTESTESHWQCITALIS--- 158  
 QY 166 GDFLVITGSLTFLSVGQCYCKRDRTEVAMIG-----VFCMLISATEITVLEARNLS 221  
 Db 159 ---AVLIHAIYT-----QCKKRCTSVITFNALPCLLAGLILSATG-WFFERQVQS 207

QY 222 SMQWSTGLLAAYVYVYALSSP--LFTLTLPFLKMSGAAP-FNLMSLTSDMWAVAIRTFIY 278  
 Db 208 TFSVSHILATLY-----LGAPAGVFGILCYFALQOKANAFQASLFLFLPIAVSLEDIY 263  
 QY 279 NOEVDWLYLAFCLVGVVGFYIYTKTEKDPNNTALENGN 317  
 Db 264 GYAISTHSMLLIPLVIGIFLTIVARNLPVTSRCRDNSS 302

RESULT 3  
 YVBV\_BACSU STANDARD; PRT; 305 AA.  
 ID YVBV\_BACSU 032256;  
 AC 032256;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HYPOTHETICAL 32.8 KDA PROTEIN IN ARAR-LACA INTERGENIC REGION.  
 GN YVBV.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC 1- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z99121; CAB15405.1; .  
 CC DR SUBTILIST; BG14088; YVBV.  
 CC DR PFAM; PF00892; DUF6; 2.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 11 31 POTENTIAL.  
 FT TRANSMEM 37 57 POTENTIAL.  
 FT TRANSMEM 70 90 POTENTIAL.  
 FT TRANSMEM 97 117 POTENTIAL.  
 FT TRANSMEM 126 146 POTENTIAL.  
 FT TRANSMEM 148 168 POTENTIAL.  
 FT TRANSMEM 180 200 POTENTIAL.  
 FT TRANSMEM 217 237 POTENTIAL.  
 FT TRANSMEM 244 264 POTENTIAL.  
 FT TRANSMEM 265 285 POTENTIAL.  
 SQ SEQUENCE 305 AA; 32820 MW; 203B2E4E7138670B CRC64;

Query Match 7.2%; Score 124.5; DB 1; Length 305;  
 Best Local Similarity 23.3%; Pred. No. 0.024;  
 Matches 74; Conservative 51; Mismatches 131; Indels 61; Gaps 13;

QY 27 LVSFDLAIM---TFSASLVSTTVDAPLTMSFTTYVVALLYGTILLYRRHKFLVPWY-- 81  
 Db 12 LLAF-LVIMGVNWPISKAALAYSPLLFAGIRTLGLGLLVIVLPRHKLRLKETWPI 70  
 QY 82 ---YALLGFDVHVG-----NVLNKAFLTSITSVSLDCWITSVIFTNWFLGTYS 132  
 Db 71 YLVSALLNTLTFGLQTIGLNLPLAGLF-----SAIVFFQFVLGVFSLNLGSEMF 122  
 QY 133 VYQFVGAACVGLLLVLLSDSGVTAAGSNPLGDFLVITGSLTFLSVGQCYCKRKO 192  
 Db 123 VMKVGILIGFAGVAVISAAGFG----GHISVIGVLLALGSAVSWALGTFYNNKTSRVD 178  
 QY 193 RIEVAM---IGVFGMLISA--TEITVLEARNALSSMOWSTGLLAAYVYVYALSSFLCTLT 247

```

Db 179 SIWVALQLTGVSFLLSIGFWTE-----SFSAIQWTAPF-----ITSLFF--IS 221
QY 248 PFLKMGGAFFNLMLTSDMAVAIRTF-----IYNQEVDMLYYLAFLVWVG 296
Db 222 VFVIALGWLFFVL-VGSGASKVASVTFLLPLISIVASSIFLHEPLTSLLAGLLIIVT 280
QY 297 IFIYTKEDPNNTNRL 313
Db 281 SICLVNTKSKAQAIAI 297

RESULT 4
YETK_BACSU
ID YETK_BACSU STANDARD; PRT; 330 AA.
AC Q31540;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 35.8 KDA PROTEIN IN LPLD-PEL INTERGENIC REGION.
GN YETK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z99107; CAB12540.1; -
DR SUBTILIST; BG12867; YETK.
DR PFAM; PF00892; DUF6; 2.
KW Hypothetical protein; Transmembrane.
FT TRANSNEM 27 47 POTENTIAL.
FT TRANSNEM 56 76 POTENTIAL.
FT TRANSNEM 90 110 POTENTIAL.
FT TRANSNEM 119 139 POTENTIAL.
FT TRANSNEM 147 167 POTENTIAL.
FT TRANSNEM 176 196 POTENTIAL.
FT TRANSNEM 206 226 POTENTIAL.
FT TRANSNEM 243 263 POTENTIAL.
FT TRANSNEM 270 290 POTENTIAL.
FT TRANSNEM 294 314 POTENTIAL.
SQ SEQUENCE 330 AA; 35758 MW; 7C6C65F1EDB276B5 CRC64;

Query Match 7.2%; Score 124; DB 1; Length 330;
Best Local Similarity 20.7%; Pred. No. 0.028;
Matches 66; Conservative 48; Mismatches 139; Indels 66; Gaps 9;

QY 20 LGGLGOLVSFDLAIMFSA-----SLVSTTVDPALPTMSFTYTV--VALLYGTILLYR 71
Db 58 LSSGLRLIASVLLMLLCIEKGFPALTKVDYFVLLVQSFTGVFLFSICLLYGV----- 112
QY 72 RHKFLPWWYVALLGVVDVHGNLYNKNKAFELTSITSVSLDCWTIVWSIIFTWNFLGTKY 131
Db 113 -----QYTGTEGILTSTPMLIGILSFFLLREKI 143
QY 132 SVYQFCAATCYGGLLLVLLSDSGVTAGSNPLLDGLFLVIT---GSILFTLSVQGEYCV 188
Db 144 EKXTLIGILLACVGMVAINLFGAGSQDGTPHALFGNNLIIAAVIGEALFTMA---KLLS 200
QY 189 KRKDRIEWMVIGVFGML-----ISATETITLERNALSSMQWSTGLLAAYVYVYVLSFLFC 244

```

```

Db 201 PHISALAISTFVSIFGLFPLFPALFEASSFDYSPVPLDWS-----YVLYYA---LFV 251
QY 245 TLFPFLKMGSG-----AAFENLSMLTSDMAVAIRTYIYNQEVDMLYYLAFLVWVGIFI 299
Db 252 TVLAFYLVWSGVTKVPAGVSGIFTSVLPVSAVLGSGVILKEPFEFVHFIGIACVIGIFV 311
QY 300 YTKTEKDPNNTRALNGL 318
Db 312 TVIKKKOPDAYPAAEKTL 330

RESULT 5
YEDA_ECOLI
ID YEDA_ECOLI STANDARD; PRT; 306 AA.
AC P09185;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 32.2 KDA PROTEIN IN DSRA-VSR INTERGENIC REGION (ORF 4).
GN YEDA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 89345179.
RA Hancock T., Gerwin N., Fritz H.-J.;
RT "Nucleotide sequence of the dcm locus of Escherichia coli K12.";
RL Nucleic Acids Res. 17:5844-5844(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12; MGI655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:11453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 97251358.
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Nakano K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasubram S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X13330; CAA31708.1; -
DR EMBL; AE000287; AAC75025.1; -
DR EMBL; D90835; CAB21710.1; -
DR PIR; J50266; J50266.
DR ECGENE; EG11141; YEDA.
DR PFAM; PF00892; DUF6; 2.
KW Hypothetical protein; Transmembrane.
FT TRANSNEM 7 27 POTENTIAL.

```



OC Actinomycetales: Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE; 98295987.  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon K.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Bigham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO ADENYL CYCLASE CLASS-3 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; Z73902; CAA98105.1; -;  
 DR TUBERCULIST; RV1318C; -;  
 DR PFAM; PF006672; DUF5; 1;  
 DR PFAM; PF00211; guanylate\_cyc; 2;  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 57 77 POTENTIAL.  
 FT TRANSMEM 90 110 POTENTIAL.  
 FT TRANSMEM 144 164 POTENTIAL.  
 FT TRANSMEM 167 187 POTENTIAL.  
 FT TRANSMEM 221 241 POTENTIAL.  
 FT TRANSMEM 257 277 POTENTIAL.  
 FT TRANSMEM 541 AA; 59331 MW; 24949CC8214386B1 CRC64;  
 SQ SEQUENCE  
 Query Match 6.4%; Score 109.5; DB 1; Length 541;  
 Best Local Similarity 22.0%; Pred. No. 0.49;  
 Matches 67; Conservative 43; Mismatches 74; Indels 121; Gaps 15;  
 QY 3 SSSRRHRENGMRT-----LLGLGLGQLVVSFDLAIMTFASLSVSTTVDAPLTMSFT 55  
 DB 40 SESQRRRRVQVMTALTAVTANLLGIVA-----LLLVTAIPEPSIVRDTPRWLTFG 93  
 QY 56 T---YTVVALLYGTLILYR-----RHKFLVPW-----YV--- 81  
 DB 94 VVPGYVLLALAGSYALTQTVQALRWAEGRKPTREERTFLAPRVAVGHLMFVGVG 153  
 QY 82 YALLGFDVHGNYLVNKAFLTSITSYILDCMTIVMSIIFTMFLGTYKSVYQVGAAI 141  
 DB 154 TALL--TTLYG--LNNAFIPREFAVSF--CGVLV---ATAVYLHTEPALRPPAAQAL 203  
 QY 142 -----CVGGLLLVLLSSDGYTAGSNPLLDGLFLVITGSILFTLSTVGQCYVK 189  
 DB 204 EAGPPRRRLAPGILGRMTVMWVLLSGV-----PVVG----- 234  
 QY 190 RKRIEVVAMIGVFGMLISATEITVLERNALSSMOWSTGLAAVYVVALSSFL-----F 243  
 DB 235 -----IALMAMFENVLL-----NLTRMQFATGVLIISNVTILVFGFILWILAW 277  
 QY 244 CTLTP 248  
 DB 278 LTATP 282  
 RESULT 8  
 Y355\_SYN3  
 ID Y355\_SYN3 STANDARD; PRT; 330 AA.

AC P74436;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HYPOTHETICAL 36.1 KDA PROTEIN SL0355.  
 GN SL0355.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE; 97061201.  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. PCC6803. II. Sequence determination of the entire  
 RT genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; D90915; BAA18537.1; -;  
 DR PFAM; PF00892; DUF6; 2;  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 15 35 POTENTIAL.  
 FT TRANSMEM 41 61 POTENTIAL.  
 FT TRANSMEM 72 92 POTENTIAL.  
 FT TRANSMEM 102 122 POTENTIAL.  
 FT TRANSMEM 125 145 POTENTIAL.  
 FT TRANSMEM 175 195 POTENTIAL.  
 FT TRANSMEM 201 221 POTENTIAL.  
 FT TRANSMEM 238 258 POTENTIAL.  
 FT TRANSMEM 264 284 POTENTIAL.  
 FT TRANSMEM 286 306 POTENTIAL.  
 FT TRANSMEM 330 AA; 36092 MW; AB44A99D8B53DAC CRC64;  
 SQ SEQUENCE  
 Query Match 6.3%; Score 108.5; DB 1; Length 330;  
 Best Local Similarity 21.5%; Pred. No. 0.36;  
 Matches 71; Conservative 59; Mismatches 141; Indels 59; Gaps 15;  
 QY 13 EGMRTLLGLGLGQLVSPDL---AIMTFASLSVSTTVDAPLTMSFTTVVVALYGTILL 69  
 DB 4 ESKTNTNIRSLGTLTAPEFLWGTAMVAMKGVLAOTTFVFAVTRILIPAGILVWAMGQK 63  
 QY 70 YRRKFLVPWYVYALLGQVVD--VHGNYLVNKAFLT--SITSVSLDCMTIVMSIIFTM 125  
 DB 64 RPQPNQOGWGIILFALVDGTLFGFLA--OGLERTAGLGSV--IIDSQPLAVALLSSWL 121  
 QY 126 F-----LGTKYSVYQVFGAAICVGLLLVLLSD-----SGVTAAGSNP 163  
 DB 122 FKEVIGGIG-----WGLLLGVGGISLIGLPDENFYQLWHLQGLSINWSG--SALGSS- 172  
 QY 164 LLGDFLVTGSLTSTLSTVGQYCVKRDRIEVVAMIGVFGMLISATBITVLERNALSSM 223  
 DB 173 --GELWMLLASLSMAVGTVLIPFVSRVRDVPVATGWMMIIGGLPLAIALVQDSEPMQNI 230  
 QY 224 Q-WSTGLLRAVYVY--ALSSFLFCTLTLPFLKMSCAAFENLSMLTS-----DMNAVAIRT 275  
 DB 231 DLWGMGNLAYATVFGSAIAYGIF-----FYLSKSG-----NLTSLSLFLTPIFALSFSN 281  
 QY 276 FYNOEVDWLYLAFCLVWVGIFVYTKTEK 305





Query Match 6.2%; Score 106; DB 1; Length 402;  
 Best Local Similarity 21.0%; Pred. No. 0.65;  
 Matches 68; Conservative 61; Mismatches 133; Indels 62; Gaps 16;

QY 11 ENEGMRRLGLGQLV-----FDLAIMTFSASLVTTVDAPLMTSFTTYVVAL--- 63  
 Db 89 EGECSGKTKVLEGLLFAMWYFNINYNKQVLK-ALHAPMTVTLVQFAGSVLITF 147  
 QY 64 YGTYLLYRRHKFLVPWYVYALLGFDVQH--GNVLNKAFLTSITSVILDCWTFIVNSII 121  
 Db 148 MWALNLYKPKISAA-QLAAILPLAVVHTLGNLTNLSGKVSFSTHTKAMEPFESV 206  
 QY 122 FTWMLGKTKSVYQVFGAAI-CVGLLLVLSDSGVTAAGSNPLLDGLFVITGSILFTLS 180  
 Db 207 LSAMFLEVPWP-VIGSLIPVGVVALASVTEFSFNWAG-----FLSAMSNLTNQS 258  
 QY 181 --TVGQCYCVKRDRIEIVAMIGVFGMLISATETIVLERNALSSMOWSTGLLAAYVYVAL 238  
 Db 259 RNVLSSKRVKMKDSDSLNITLFSI-----ITLMSL-----FLMAPVTFFS 298  
 QY 239 SSFLFCTILTPLLKMSGAFFNLSSMLTSDMWAVAIRFIYNQVDWL-----YVLA 289  
 Db 299 EGIRF---TPSYIQSAGV---NQOQIYTKSLIALCFHAY-QQVSYMLARVSPVTHSVG 351  
 QY 290 FC-----LVWVGIFTYTKTKDPNN 309  
 Db 352 NCVRVVIVSSVIFTKTPSPVN 375

RESULT 11  
 YBFH\_BACSU STANDARD; PRT; 306 AA.  
 AC Q31448; 1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HYPOTHETICAL 33.8 KDA PROTEIN IN GLPT-PURT INTERGENIC REGION.  
 GN YBFH.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RA Haga K., Liu H., Yasumoto K., Takahashi H., Yoshikawa H.;  
 RT "Sequence analysis of the 70Kb region between 17 and 23 degree of the  
 RT Bacillus subtilis chromosome."  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB006424; BAA33118.1; -  
 DR EMBL; 299105; CAB12015.1; -  
 DR SUBTILIST; BG12737; YBFH.  
 DR PFAM; PF00892; DUF6; 2.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 11 31 POTENTIAL.  
 FT TRANSMEM 34 54 POTENTIAL.  
 FT TRANSMEM 68 88 POTENTIAL.  
 FT TRANSMEM 97 117 POTENTIAL.  
 FT TRANSMEM 123 143 POTENTIAL.  
 FT TRANSMEM 151 171 POTENTIAL.  
 FT TRANSMEM 188 208 POTENTIAL.  
 FT TRANSMEM 222 242 POTENTIAL.

FT TRANSMEM 248 268 POTENTIAL.  
 FT TRANSMEM 271 291 POTENTIAL.  
 SQ SEQUENCE 306 AA; 33842 MW; 7DE953E93B2FB072 CRC64;

Query Match 6.1%; Score 105; DB 1; Length 306;  
 Best Local Similarity 21.2%; Pred. No. 0.59;  
 Matches 64; Conservative 63; Mismatches 143; Indels 32; Gaps 14;

QY 32 LAIMTFSASLVSTVDAPLMTSFTTYV--ALLYGTYLLYRRHKFLVPW-YW-----YA 83  
 Db 15 VTILWGTTFISTKRV---LLADFSPMEILFYFLMGIALILVPRNIPFERNWQELFA 71  
 QY 84 LLGVDVGVNVLV-NKAFELTSITSVS-ILDCWTFIVNSIIFTWMLGTYKYVYQVFGAAI 141  
 Db 72 GAGLFGVTLFLEENIALTYASNVGMIVSIIPMTAVLAHFLLEGEKLELTFLIGFIS 131  
 QY 142 CVGGLLLVLSDSGVTAAGSNPL-----LGDFLVITGSILFTLSVVGQCYCVKRDRIEIV 196  
 Db 132 ALIGLLITF--NGNVVLRNPLGDINAAGAALVFGVGSIFPMKKLSAYEYHI-----IEL 184  
 QY 197 VAMIGVFGMLISATEITVLEARN-ALSSMOWSTGLL-RAVYVVALSSFLFCTILTPFLKMS 254  
 Db 185 TQRFVLYGLLFMPALFLDFHFDLSRFSSASNLNMLFLGIGASALCFATWN-YVGVVL 243  
 QY 255 GAAPFNLSMLTSDMWAVAIRFIYNQVDWLYLAFLGVVVGIFITYTKTKDPNTRALE 314  
 Db 244 GAVKSSAYIYVNPVITTAASVLIHENMTIALGGALTLLGLYI---SEUKP-KAKLLE 299  
 QY 315 NG 316  
 Db 300 NG 301

RESULT 12  
 COX1\_TRYBB STANDARD; PRT; 549 AA.  
 ID COX1\_TRYBB  
 AC P04371;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DE 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 OS COXI OR COI.  
 OS Trypanosoma brucei.  
 OS Mitochondrion.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85037915.  
 RA Hensgens L.A.M., Brakenhoff J., de Vries B.F., Sloof P., Tromp M.C.,  
 RA van Boom J.H., Benne R.;  
 RT "The sequence of the gene for cytochrome c oxidase subunit I, a  
 RT frameshift containing gene for cytochrome c oxidase subunit II and  
 RT seven unassigned reading frames in Trypanosoma brucei mitochondrial  
 RT maxi-circle DNA."  
 RL Nucleic Acids Res. 12:7327-7344(1984).  
 RN [2]  
 RP SEQUENCE OF 1-402 FROM N.A.  
 RC STRAIN-164;  
 RX MEDLINE; 85240413.  
 RA Payne M., Rothwell V., Jasmer D.P., Feagin J.E., Stuart K.;  
 RT "Identification of mitochondrial genes in Trypanosoma brucei and  
 RT homology to cytochrome c oxidase II in two different reading  
 RT frames";  
 RL Mol. Biochem. Parasitol. 15:159-170(1985).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B.  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) - 2 H(2)O +

```

CC 4 FERRICYTOCHROME C.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
CC EMBL: X01094; -; NOT_ANNOTATED_CDS.
CC EMBL: M14820; AAB59223.1; -.
CC PIR: A00471; ODU7MB.
CC HSP: P00396; LOCC.
CC PFAM: PF00115; COX1; 1.
CC PRINTS: PR01165; CYCOXIDASE1.
CC PROSITE: PS00077; COX1; 1.
CC Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 64 64 IRON (HEME A) (PROBABLE).
FT METAL 243 243 COPPER B (PROBABLE).
FT METAL 247 247 COPPER B (PROBABLE).
FT METAL 292 292 COPPER B (PROBABLE).
FT METAL 293 293 COPPER B (PROBABLE).
FT METAL 378 378 IRON (HEME A3) (PROBABLE).
FT METAL 380 380 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 549 AA; 63348 MW; 966050DDB7FFC2F0 CRG64;

Query Match 6.1%; Score 105; DB 1; Length 549;
Best Local Similarity 21.0%; Pred. No. 1;
Matches 64; Conservative 51; Mismatches 128; Indels 62; Gaps 13;

QY 22 LGGLGVLPDGLALMTFS-----ASLVSTTVTDAPLTWSTFTYTVVALLYGTYLLY 70
Db 226 VGGDLVLVQHLFPFGHPGHEVYIIILPVLGYSTIE--VTSPRCVFSVAMIYMLLIS 283
QY 71 RRHKFLVPWYVYVALLGVDFVHGVNLYNKAFLTSITSVSLDCWTIVWSIIFFWFLGPK 130
Db 284 VLGMFV--WAHMFVVGMDVDSRAYFGSITVLIGLPTCKLFNW--IYSFLFTDNCICE 339
QY 131 -YSYGVFGVCAICVGLLVLVLLSDSCVTAAGSNPLGDLFLVITGSLTFLSTVGYQYCVK 189
Db 340 IYFIYMFV-LMFLAGLGLTGLFSNVGI-----DILMHDYFVVAHPHYVLS----- 384
QY 190 RKDRIEVVAMIGVFG-----MLISATEITVLERNALSSNQWSTGLLAAYVYVYVYVLSFLF 243
Db 385 -----LGAVGVGVGGFFHFLMKWIFELHTLWLFFFISTLWFGSNWFFPLSLGMPAF 438
QY 244 CTLTPFLKMGAGAFNLNMLTSDMVAIRTF-----IYNQYVDWLYYLAFCILVYVG 296
Db 439 ----PRISDYPISELWSAFT--LYGMLLLTFLVFCCLPNVILFDY----CLFFIN 488
QY 297 IFIYT 301
Db 489 LFTYS 493

RESULT 13
TSCC_HUMAN
ID TSCC_HUMAN STANDARD; PRT; 1021 AA.
AC P5017;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE THIAZIDE-SENSITIVE SODIUM-CHLORIDE COTRANSPORTER (NA-CL SYMPORTER).
GN SLC12A3 OR TSC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS GS.
RX MEDLINE: 96122035.
RA Simon D.B., Nelson-Williams C., Bia M.J., Ellison D., Karet F.E.,
RA Molina A.M., Vaara I., Iwata F., Cushner H.M., Koolen M., Gaiza F.J.,
RA Gitelman H.J., Lifton R.P.;
RT 'Gitelman's variant of Bartter's syndrome, inherited hypokalemic
RT alkalosis, is caused by mutations in the thiazide-sensitive Na-Cl
RT cotransporter.';
RL Nat. Genet. 12:24-30(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Mastrolanni N., de Fusco M., Zollo M., Arrigo G., Zuffardi O.,
RA Bettinelli A., Ballabio A., Casari G.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC -1- FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A
CC MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PREDOMINANT IN KIDNEY.
CC -1- DISEASE: DEFECTS IN SLC12A3 ARE THE CAUSE OF GITELMAN'S SYNDROME
CC (GS), AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY DIVERSE
CC ABNORMALITIES IN ELECTROLYTE HOMEOSTASIS INCLUDING HYPOKALAEMIC
CC METABOLIC ALKALOSIS. GS IS A SUBSET OF BARTTER'S SYNDROME.
CC -1- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
CC EMBL: U44128; AAC50355.1; -.
CC EMBL: X91220; CAA62613.1; -.
CC MIM: 600968; -.
CC MIM: 263800; -.
KW Transport; Transmembrane; Glycoprotein; Disease mutation.
FT DOMAIN 1 135 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 136 156 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 157 158 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT DOMAIN 180 218 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 219 229 POTENTIAL.
FT DOMAIN 240 261 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 262 282 POTENTIAL.
FT DOMAIN 283 286 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 287 307 POTENTIAL.
FT DOMAIN 308 339 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 340 360 POTENTIAL.
FT DOMAIN 361 377 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 378 398 POTENTIAL.
FT TRANSMEM 399 452 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 453 473 POTENTIAL.
FT DOMAIN 474 511 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 512 532 POTENTIAL.
FT DOMAIN 533 555 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 556 577 POTENTIAL.
FT DOMAIN 578 598 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 599 660 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 661 681 POTENTIAL.
FT DOMAIN 682 1021 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 406 406 POTENTIAL.
FT CARBOHYD 426 426 POTENTIAL.
FT VARIANT 209 209 R -> W (IN GS).
FT VARIANT 349 349 P -> L (IN GS).
FT VARIANT 421 421 C -> R (IN GS).
FT VARIANT 421 421 /FTID-VAR_007115.
FT VARIANT 421 421 /FTID-VAR_007114.
FT VARIANT 421 421 /FTID-VAR_007115.

```

```
FT VARIANT 486 486 D -> N (IN GS).
FT VARIANT 496 496 /FTid-VAR_007116.
FT VARIANT 561 561 G -> C (IN GS).
FT VARIANT 561 561 /FTid-VAR_007117.
FT VARIANT 588 588 MISSING (IN GS).
FT VARIANT 588 588 /FTid-VAR_007118.
FT VARIANT 630 630 A -> V (IN GS).
FT VARIANT 630 630 /FTid-VAR_007119.
FT VARIANT 655 655 G -> V (IN GS).
FT VARIANT 655 655 /FTid-VAR_007120.
FT VARIANT 655 655 R -> H (IN GS).
FT VARIANT 655 655 /FTid-VAR_007121.
FT VARIANT 728 728 R -> L (IN GS).
FT VARIANT 728 728 /FTid-VAR_007122.
FT VARIANT 741 741 A -> T (IN GS).
FT VARIANT 741 741 /FTid-VAR_007123.
FT VARIANT 850 850 G -> R (IN GS).
FT VARIANT 850 850 /FTid-VAR_007124.
FT VARIANT 955 955 L -> P (IN GS).
FT VARIANT 955 955 /FTid-VAR_007125.
FT VARIANT 955 955 R -> Q (IN GS).
FT VARIANT 955 955 /FTid-VAR_007126.
FT CONFLICT 459 460 AG -> VV (IN REF. 2).
FT CONFLICT 766 766 E -> D (IN REF. 2).
FT CONFLICT 807 807 V -> GARPSVSGAL (IN AAC50355).
SQ SEQUENCE 1021 AA; 113138 MW; D7ECE53DA6233821 CRC64;

Query Match 6.1%; Score 105; DB 1; Length 1021;
Best Local Similarity 21.3%; Pred. No. 1.9;
Matches 44; Conservative 33; Mismatches 78; Indels 52; Gaps 9;

QY 110 ILDCWTV-----W-----SIIFTWFLGTYKSYVGFVGAICV-----GGLLLV 149
DB 147 MLNIWGVILYLRPWITAQAGIVLTWIIILSVTVTSIGLSISAISTNKVKSGGYFL 206
QY 150 LLSDSGVTAAGSNPLLDGLFVITGSILFTLSTVG-----QEYCVKRRDREVVAMI 200
DB 207 ISRSLGPELGS---IGLIFANAVGVAMHTVGFETVRDLQEGAPIVDPIINDIRII 263
QY 201 GVFGMLISATEITVLERNALSSQWSTG---LLAAYVYALSSFLFTCLPFLFLKMSGAA 257
DB 264 GV-----VSVTVLLAISLAGMEWESKAQVLFVLVMSFANYLVGTLIPPSDEKASKG 316
QY 258 FFLNLSMLTSDMWAVAIPTFIYNOEVDW 284
DB 317 FFS---YRADI-----FVONLVPDW 333

RESULT 14
YCAL_CLOKL STANDARD; PRT; 311 AA.
AC P38943;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 33.5 KDA PROTEIN IN CAT1 5'REGION (ORFY).
OS Clostridium kluyveri.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 555;
RX MEDLINE; 96146540.
RA Soehling B., Gottschalk G.;
RT "Molecular analysis of the anaerobic succinate degradation pathway in
RT Clostridium kluyveri."
RL J. Bacteriol. 178:871-880(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE UPF0077 (PSCM) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensed@isb-sib.ch).
CC -----
DR EMBL; L21902; AAA92345.1; -
DR PFAM; PF00892; DUF6; 2.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 97 117 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 185 203 POTENTIAL.
FT TRANSMEM 219 239 POTENTIAL.
FT TRANSMEM 244 264 POTENTIAL.
FT TRANSMEM 265 285 POTENTIAL.
SQ SEQUENCE 311 AA; 33507 MW; A5E7CBE9EC9ADF0B CRC64;

Query Match 6.0%; Score 103.5; DB 1; Length 311;
Best Local Similarity 19.1%; Pred. No. 0.76;
Matches 48; Conservative 51; Mismatches 121; Indels 31; Gaps 8;

QY 81 WYALLGEVDVH-GNYLVNKAPELTISVSILDCWTVWTSIIFTWFLGTYKSYVQFVGA 139
DB 70 YLALCGLIVASISMSILOFVYTKASTAAVLCFNAFTTIPFAIFLKEKIKGITVSI 129
QY 140 AICVGGLLV-----LLSDSGVTAAGSNPLLDGLFVITGSILFTLSTVGQYCVKRRDRI 194
DB 130 IVSLIGVVIIFNPAKVMEGIG---GSRDLIGICFALVAAVVWSL-----YTVISKRI 179
QY 195 EVA-----MIGVGMILISATEITVLERNALSSQWSTGLLAAYV-VYALSSFLFC 244
DB 180 EIYGVGVVENCISFFFGVIALLL---LVVTGRPIFGITLNLNLLVLLYMGFIKAVGYIC 236
QY 245 TLFPFLKMSGAAFFNLMSLTSDMWAVAIPTFIYNOEVDWLYLAFCLVTVVGIYIKTE 304
DB 237 YLG--AKETSATVASTVFLIKPALATVLAILLGESIEVNVVIGVFIIGSIINSSN 294
QY 305 KDPNTRALEN 315
DB 295 KKANDLKKVAN 305

RESULT 15
YMB8_YEAST STANDARD; PRT; 393 AA.
AC Q03730;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 43.7 KDA PROTEIN IN OST6-PSP2 INTERGENIC REGION.
GN YML018C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO YEAST YDR438W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; 246659; CAA86637.1; -.
KW Hypothetical protein; Transmembrane
FT TRANSMEM 18 38 POTENTIAL.
FT TRANSMEM 47 67 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 178 198 POTENTIAL.
FT TRANSMEM 220 240 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT TRANSMEM 322 342 POTENTIAL.
FT TRANSMEM 346 366 POTENTIAL.
SQ SEQUENCE 393 AA; 43748 MW; DEC5C3EEF293F5ED CRC64;

Query Match          6.0%; Score 103.5; DB 1; Length 393;
Best Local Similarity 21.9%; Pred. No. 0.96;
Matches 63; Conservative 46; Mismatches 122; Indels 57; Gaps 13;

QY 63 LYGTILLYRRHKFLPYWYVYALLGFVDVHGNYLVNKAPELTSITSVILDCWTIYWSIIF 122
DB 123 LYETIKL--SAEFCILWF-----TANLVTNASLAFTSVASQTILS-----TTSSEF 166
QY 123 TWMLGT-----KSYVYQFVGAACVGGLLLVLLSD-----SGVTAAGSNPLL 165
DB 167 T-LFIGAICHVESLSKSVLGSEIFSVGIIMWTKSDSHORYORHIAVSGDDNDNAVQVLI 225
QY 166 GDFLVITGSILF-TLST-----VGQECYCKRKDRIEVAMIGVFGMLISATEITVLERNA 219
DB 226 GNLLALAGAVLYGVYSTLLKREVGDETRVNMK--IFFGFVGLFNLFLWPSLIVLDFFG 282
QY 220 LS--SMQSTGLLAAYVYVALSSFL--FCTLTPFLKMSGAFFNLSMLTSDMNAVAIRT 275
DB 283 WEPFSLPKDPKVYVVIIFVNCILTFVSDFCWAKAMLTSPLTVTVGLSI-----TIPLAM 336
QY 276 F----IYNQVDWLYLAFCLVGVVIGFIVTKTEKDPNTRALENGNLD 319
DB 337 FGDVIFKHTMSALYLFGATLILGSEFIINKSSEEHFENSITASNYE 384
```

Search completed: August 1, 2000, 21:32:27  
Job time: 6047 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:27:47 ; Search time 51.03 Seconds  
(without alignments)  
406.259 Million cell updates/sec

Title: US-09-147-955-10  
Perfect score: 1723  
Sequence: 1 M5SSSRWRWRENGMRRTLL.....GNLDHEYSLEDDQDDTPRPK 335

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_64 : \*  
1: pir1 : \*  
2: pir2 : \*  
3: pir3 : \*  
4: pir4 : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	428	24.8	505	2	T41303	hypothetical prote
2	145	8.4	324	2	S62543	hypothetical prote
3	134.5	7.8	292	2	D71181	hypothetical prote
4	125.5	7.3	370	2	S69718	hypothetical prote
5	124.5	7.2	305	2	F70030	conserved hypothet
6	124	7.2	330	1	H69798	conserved hypothet
7	122.5	7.1	341	2	T40424	hypothetical prote
8	121	7.0	306	2	J50266	membrane protein y
9	120.5	7.0	411	2	T03836	phosphate/phosphoe
10	119	6.9	490	2	A46391	CAMP receptor subt
11	114	6.6	311	2	A72487	hypothetical prote
12	114	6.6	567	2	S58750	NADH dehydrogenase
13	112.5	6.5	294	2	C72378	sugar ABC transpor
14	112.5	6.4	267	2	C75131	hypothetical prote
15	110.5	6.4	340	2	T01949	hypothetical prote
16	110.5	6.4	445	2	D69902	sodium-dependent t
17	109.5	6.4	541	2	B70769	hypothetical prote
18	109	6.3	442	2	S50332	NADH dehydrogenase
19	108.5	6.3	330	2	S76408	hypothetical prote
20	108.5	6.3	383	2	T38194	hypothetical prote
21	107	6.2	618	2	G75113	NADH dehydrogenase
22	106.5	6.2	531	2	S30011	probable membrane
23	106	6.2	390	2	T04096	glucose-6-phosphat
24	106	6.2	402	2	T14438	phosphate/triose-p
25	105.5	6.1	378	2	G70435	hypothetical prote
26	105.5	6.1	434	2	T40700	phosphate-phosphoe
27	105	6.1	306	2	C69749	conserved hypothet
28	105	6.1	549	1	ODUTMB	cytochrome-c oxida
29	105	6.1	617	2	F71019	hypothetical prote

# ALIGNMENTS

## RESULT 1

T41303

hypothetical protein SPCC320.08 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 04-Mar-2000

C:Accession: T41303

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Wedler, H.; Wambutt, R.; Wedler, E.

submitted to the EMBL Data Library, March 1998

A:Reference number: Z21986

A:Accession: T41303

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-505 <WOO>

A:Cross-references: EMBL:AL022245; PIDN:CAA18310.1; GSPDB:GN00068; SPDB:SPCC320.08

A:Experimental source: strain 972h-; cosmid c320

C:Genetics:

A:Gene: SPDB:SPCC320.08

A:Map position: 3

C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPCC320.08

Query Match	24.8%	Score 428;	DB 2;	Length 505;
Best Local Similarity	31.4%	Pred. No. 3e-27;		
Matches 93;	Conservative 65;	Mismatches 108;	Indels 30;	Gaps 7;
Qy	24	LGQLVSFDLAIMTFSASLVSTTVDAPLTMSFTTYVVALYGTILLYR-----R	72	
Db	147	LGQVLSLCITATNFGYMSGISNIPAFQTLVYALLTLVTPYVFRMGFKKYEIMFR	206	
Qy	73	HKFLVPWYV-YALLGFVDVHGNYLNKAFELTSITSVSLDCWTIVWSIIFTWMLGKY	131	
Db	207	HG-----WKYIIFFAFVDEGNYFVLAYQYTNMLUSALLDSMATVAVVILSFILVKRY	260	
Qy	132	SVYQFVGAACVGGILLVLLDSGVTAAGS----NPLGLDFLVIITGTSILFTLSTGQYVC	187	
Db	261	HWSQILGVVACIGLVLLVSD--VISRGDYSANVPGLDGYMIIGATCYGVSNLLEYF	318	
Qy	188	VKRDRIEVAMIGVFGMLISATEITVLERNALSSMONST---GLLAAYVVALSFLFC	244	
Db	319	ASKLPLYVVIQGLSLYSGSIISIQTFIDRRHLYTLHWTSEMGGYLAGFI---LVMLFLY	375	
Qy	245	TLTFPLKMSGAFFNLNLSLSDMAVAIRFIYQNEVDWLYLAFCLVVGIFTY	300	
Db	376	SLAPILFRMSSATFYNISLLTSDFSLVIGHFVGYHWLYPIAFVLIILGLFY	431	

## RESULT 2

S62543

hypothetical protein SPAC12G12.12 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Jan-2000

C:Accession: S62543; T37589

R:Devlin, K.; Odell, C.; Churcher, C.M.  
 submitted to the EMBL Data Library, November 1995  
 A:Reference number: S62532  
 A:Accession: S62543  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-324 <DEV>  
 A:Cross-references: EMBL:Z66568; NID:g1052518; PID:CAA91507.1; PID:g1052530  
 R:Devlin, K.; Odell, C.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
 submitted to the EMBL Data Library, November 1995  
 A:Reference number: Z21727  
 A:Accession: T37589  
 A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-324 <DE2>  
 A:Cross-references: EMBL:Z66568; PID:CAA91507.1; GSPDB:GN00066; SPDB:SPAC12G12.12  
 A:Experimental source: strain 972h; cosmid c12G12  
 C:Genetics:  
 A:Gene: SPAC12G12.12  
 A:Map position: 1L  
 A:Introns: 19/2; 254/2

Query Match 8.4%; Score 145; DB 2; Length 324;  
 Best Local Similarity 23.8%; Pred. No. 0.00014;  
 Matches 67; Conservative 42; Mismatches 117; Indels 56; Gaps 10;  
 QY 68 LLYRRHKFLYPWY-----WYALLGFVDVHGNVYNKAFELTSITSVILDCWT 115  
 Db 53 VLYKRHS-QGPGESDHLPLKHKFMALPAIMDIOGSLMNVGLLYTSASIVQMRGSL 111  
 QY 116 IVMSIIFTMFLGTYKVYQFVGAACVGGLLVLLSDSGVTAAGSNPLGDFLVTGSI 175  
 Db 112 IIFVALPATLLARTICQLQWLSLFWLGVAVIGVSGSS-SSIGSNPIIGITAVLIGQF 170  
 QY 176 LFTLSTVGQBYCVK--RKDRIVVAMTGVFGLISATEITVLERNALSSMOWSTGLLAA 233  
 Db 171 FLATQFTIEEYILSFQVDPSELVAVEGTYGVFVLGMII-----SYFFIGSTTAY 223  
 QY 234 -----VYALSSFLCFLTPFLKMSCA----AFPNLSML-----T 265  
 Db 224 HGWFDSHVISR-----NEVPALVVISGVILVSIAPFNVSLATKLSATRSLLDIART 280  
 QY 266 SDMWAVARTFIYNQEDVLYLAFCLVYVVGIFYTKTERDP 307  
 Db 281 FGIWIIAMANG--ESPHLLQFLGVLLIYGIYHSIIKFP 320

RESULT 3  
 D71181  
 hypothetical protein PH1729 - Pyrococcus horikoshii  
 C:Species: Pyrococcus horikoshii  
 C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 14-Aug-1998  
 C:Accession: D71181  
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekir  
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
 DNA Res. 5, 55-76, 1998  
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
 A:Reference number: A71000; MUID:98344137  
 A:Accession: D71181  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-292 <KAW>  
 A:Cross-references: GB:AP000007; NID:g3236134; PID:d1031786; PID:g3258160  
 A:Experimental source: strain OT3  
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C:Genetics:  
 A:Gene: PH1729

Query Match 7.8%; Score 134.5; DB 2; Length 292;  
 Best Local Similarity 22.0%; Pred. No. 0.0009;  
 Matches 45; Conservative 51; Mismatches 80; Indels 29; Gaps 8;

QY 20 LGLGLGOLVSFDLAI---MTFSASLVSTTVDAPLMTSFTTYYVVALYGTILYRRHKEL 76  
 Db 6 MNFALGVILAFSAFTWALTSVLKSVSMRRVSPDLTNFLRLVIASGFYLLIYLN---L 62  
 QY 77 VP-----WY-----WYALLGFVDVHGNVYNKAFELTSITSVILDCWTIYMSIIFTMMFL 127  
 Db 63 IPSKGIMWVIVVFGSIIGF--NIGDWLFLEGKMLLGVSRANMLTTLHPITWVLAHYLL 120  
 QY 128 GTYSVYQFVGAACVGGLLVLLSDS---GVTAAGSNPLGDFLVTGSIILFTLSTVGQ 184  
 Db 121 GRPLNAYFLGAGLIILAVLLLSKSESGGINARGV-----LLVVVAELLTIALIT 174  
 QY 185 EYCVKRRDRIEVAM---IGVFGML 206  
 Db 175 DMLVSNPPVPIIAALRISSGILGGL 199

RESULT 4  
 S69718  
 hypothetical protein YDR438w - yeast (Saccharomyces cerevisiae)  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 29-Oct-1999  
 C:Accession: S69718  
 R:Dieckrich, F.S.  
 submitted to the EMBL Data Library, August 1995  
 A:Description: The sequence of S. cerevisiae lambda 3641 and cosmids 9461, 9831, and  
 A:Reference number: S69553  
 A:Accession: S69718  
 A:Molecule type: DNA  
 A:Residues: 1-370 <DIE>  
 A:Cross-references: EMBL:U33007; NID:g927685; PIDN:AA64860.1; PID:g927694; GSPDB:GNO  
 C:Genetics:  
 A:Gene: MIPS:YDR438w  
 A:Map position: 4R  
 C:Superfamily: probable membrane protein YMD018c

Query Match 7.3%; Score 125.5; DB 2; Length 370;  
 Best Local Similarity 21.8%; Pred. No. 0.0062;  
 Matches 57; Conservative 57; Mismatches 102; Indels 45; Gaps 13;  
 QY 82 YALLGFVDVHGNVYNKAFELTSITSVILDCWTIYMSIIFTMMFLGTR-----YSYQF 136  
 Db 124 FCVLWFV---ANLAANAALSYTTIVASSTILSS---TSSEFT-LFLATSLGIEFTSKKL 175  
 QY 137 VGAACVGGLLVLLSDSGV--TAAGSNPLGDFLVTGSIILFTLSTVGQYCVKRD-R 193  
 Db 176 LGLFVSLFGIILIVMQSKQODSVSASSFLVGNLTALLGSLGYSVYTTLLKYEISSRGLR 235  
 QY 194 IEVYAMIGVFQM-----LISATEITVLERNAL-SSMOWSTGLLAAVYVVALSSFLFC 244  
 Db 236 LDIQMFLGYGVGIFTFLFWPILIIDITHMETPELPSNFHISFLVMUNCIIFVSDIFWC 295  
 QY 245 ----TLTPFLKMSGAFFNLMLSDMVAIVAIKFTYINQEDVLYLAFCLVYVVGIFY 300  
 Db 296 KALLTSLPVYVVALTFTIPLAFADPVRAREFT-----PW-YIIGVIFIVSFVLV 347  
 QY 301 -----TKTERDPNNTALENG 316  
 Db 348 NHRGESAVEKD---CAAVEKG 365

RESULT 5  
 F70030  
 conserved hypothetical protein ywbv - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 C:Accession: F70030  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
 A.; Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
 Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler  
 Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zunstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: F70030  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-305 <KUN>  
 A:Cross-references: GB:299121; GB:AL009126; NID:g2635827; PIDN:CAB15405.1; PID:ell186088;  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yvbY

Query Match 7.28; Score 124.5; DB 2; Length 305;  
 Best Local Similarity 23.38; Pred. No. 0.0061;  
 Matches 74; Conservative 51; Mismatches 131; Indels 61; Gaps 13;  
 QY 27 LVSPDLAIM---TFSASIVSTVDAPLMTSPTTVVVALYGTLLVYRRHKFLVPWY-- 81  
 DB 12 LQAP-LVIMGVNPLSKAALAYSPPLFAGIRTLGGLLVIVVALPRIHKLKURKETWPI 70  
 QY 82 ---YLLGFVDVHG-----NYLVNKAFLTSITSVSLDCTWIVWSIIITFMFLGTRYS 132  
 DB 71 YLVSALLNITFLYGLQITGLNYPAGLF-----SAIVFFQFVLMGVSNLWGESMF 122  
 QY 133 VYQFVGAACVGGLLVLLSDSGVTAGSNPLLDGFLVIGTSILFTLSTVGQYCVARKD 192  
 DB 123 VMKVIIGLIFAGVAVISAAGFG----GHISVIGVLLALGSAVSWALGTVMKKTGRVD 178  
 QY 193 RIEVAM---IGVFGMLISA--TEITVLERNALSSMOWSTGLLAAYVYVALSFLFCTLT 247  
 DB 179 SIWVVALQITGVSFLLISGWTE-----SFAIQWTAPF-----ITSLLF--IS 221  
 QY 248 PFLKMGAAFFNLSMLTSDMVAIVTF-----IYNQEVDMVLYAFCLVWVG 296  
 DB 222 VFVIALGWLVEFTL-VGSGEASKVASTFLPLISIVASSIFLHEPTLSLGLLIIVT 280  
 QY 297 IFITKTEKDPNNTRAL 313  
 DB 281 SICLVNTRKSAQKAAAI 297

RESULT 6  
 H69798  
 conserved hypothetical protein yekK - *Bacillus subtilis*  
 C:Species: *Bacillus subtilis*  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: H69798  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
 C.; Bron, S.; Brouillet, S.; Bruchsi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler  
 Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zunstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033

A:Accession: H69798  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-330 <KUN>  
 A:Cross-references: GB:299107; GB:AL009126; NID:g2632866; PID:ell182700; PID:g2633034  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yekK  
 C:Superfamily: hypothetical protein yedD  
 Query Match 7.28; Score 124; DB 1; Length 330;  
 Best Local Similarity 20.78; Pred. No. 0.0073;  
 Matches 66; Conservative 48; Mismatches 139; Indels 66; Gaps 9;  
 QY 20 LGLGIGOLVFEFDLAIMFSA-----SLVSTTVVDAPLMTSPTTVV--VALLYGTILLYR 71  
 DB 58 LSSGLRFLIASVWLLMLLFCIEKGFPALTKKDVFLVLSQSFVTGFLFSICLLYGV---- 112  
 QY 72 RHKFLVPWYVYALLGFDVHGNYLVNKAFLTSITSVSLDCTWIVWSIIITFMFLGTRYS 131  
 DB 113 -----QYTTGESCILTSTTPMLIGILSFFLLREKI 143  
 QY 132 SVYQFVGAACVGGLLVLLSDSGVTAGSNPLLDGFLVIT---GSILFTLSTVGQYCV 188  
 DB 144 EKKTLIGILLAVCGVMAINLFGAGSQDGTPHALFGNMLIIAIVIGALFTLMA---KLLS 200  
 QY 189 KRKRIVVAMIGVFGML-----ISATEITVLERNALSSMOWSTGLLAAYVYVALSFLFC 244  
 DB 201 PHISALAISTFVSFLGFLFLFPALFEASSPDYVPTVLDWS-----YVLYA---LDF 251  
 QY 245 TLTPFLKMGSG-----AAFFNLMLTSDMVAIVTFITVNOEVDWLYVLAFLCVVVGFI 299  
 DB 252 TVLAFYLMYSCVTVPACVSGIFTSVLPVSAVLSGVILKEPFEFVHFIGACVIGGIFV 311  
 QY 300 YTKTEKDPNNTRALENGL 318  
 DB 312 TVIKKQPDAPYAAAEKTL 330

RESULT 7  
 T40424  
 hypothetical protein SPBC405.03c - fission yeast (*Schizosaccharomyces pombe*)  
 C:Species: *Schizosaccharomyces pombe*  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
 C:Accession: T40424  
 R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, March 1999  
 A:Reference number: Z21928  
 A:Accession: T40424  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-341 <SEE>  
 A:Cross-references: EMBL:AL035655; PIDN:CAB38602.1; GSPDB:GN000067; SPDB:SPBC405.03c  
 A:Experimental source: strain 972h-; cosmid c405  
 C:Genetics:  
 A:Gene: SPDB:SPBC405.03c  
 A:Map position: 2  
 C:Superfamily: probable membrane protein YML018c

Query Match 7.18; Score 122.5; DB 2; Length 341;  
 Best Local Similarity 19.08; Pred. No. 0.0099;  
 Matches 64; Conservative 65; Mismatches 121; Indels 87; Gaps 13;  
 QY 27 LVSPDLAIMTFSASIVSTVDAPLMTSPTTVVVALYGTLLVYRRHKFLVPWY-- 81  
 DB 15 LLLFWFLWLISSELTSSLLDD--DNFSPFLIYINTGTFFVY-----LIPWFSEKKT 67  
 QY 82 -----YALLGFDV--HGNYLVNKAFL 103  
 DB 68 RKHRLMSELSMYESVHDSFNLGTRPNPLGRQTRAYLSLGFCTIIFAAVNFSSSLGFT 127





```

Query Match          6.98; Score 119; DB 2; Length 490;
Best Local Similarity 22.3%; Pred. No. 0.028;
Matches 79; Conservative 47; Mismatches 116; Indels 112; Gaps 21;

QY 18 TLGLG--LGLVSEDLAINTF--SASLVSTTVDAPLTM-----SETTYTVVALLY 64
    || || || || || || || || || || || || || || || || || || || ||
Db 39 TLVLGFWRLKLLRNHITKIITPCFSTSLAKDLSTILTLEKOSNGSPCY-----LY 93
    || || || || || || || || || || || || || || || || || || || ||
QY 65 GTILYRRHKFLPWPYWYALGFDVHGNYLVNKAFLTSTSILDCWRI--VWSIIF 122
    || || || || || || || || || || || || || || || || || || || ||
Db 94 ATVITYGS--LACWLMTLCLSF-SIV-NLIVKREPEPEKEKYHYHFCWVPFMSVIM 148
    || || || || || || || || || || || || || || || || || || || ||
QY 123 -----TWMFLGPKYSYQ---FVGAICVGGLLLVLLSDSGVTAAGSNPLLGDFL 169
    || || || || || || || || || || || || || || || || || || || ||
Db 149 LSKGVIEVTGNCWIGNTYGVYRGFLVGPFLAIWFLAALV--GLTSRYT-----YK 199
    || || || || || || || || || || || || || || || || || || || ||
QY 170 VITGSILFTLSTVGQGYCVKKRDR-----IEWAMIGVFGMLISATEITVLER--NALSS 222
    || || || || || || || || || || || || || || || || || || || ||
Db 200 VIRSV-----SNNKDRHMYQFKLINKIIVF-----LICWFFAVINRVINGLNM 244
    || || || || || || || || || || || || || || || || || || || ||
QY 223 MQNSTGLLAAYVYVALSSFLCTLTPLFLKMGAAFNLSMLTSDMMAVAIRTFYNOEV 282
    || || || || || || || || || || || || || || || || || || || ||
Db 245 FPAWVSILHTY-----LSVSHGFYASV-----TFIYNPL 274
    || || || || || || || || || || || || || || || || || || || ||
QY 283 DWLYYLAFLV--VVGIFYT--KTEKDPNN-----TRALENGMLDHEYS 323
    || || || || || || || || || || || || || || || || || || || ||
Db 275 MWRYLASILIPFTKFGYFVETQORLEKKNNNHNSHPVGLSNNAQNHHHHHN 328
    || || || || || || || || || || || || || || || || || || || ||

RESULT 11
A72487
hypothetical protein APE2537 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72487
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no,
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazac
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon
A:Reference number: A72450; MUID:99310339
A:Accession: A72487
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-311 <RAW>
A:Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAAB1553.1; PID:d1045339;
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE2537

```

Query Match	6.6%	Score 114;	DB 2;	Length 311;
Best Local Similarity	20.3%	Pred. No. 0.044;		
Matches	62;	Conservative	51;	Mismatches 127;
			Indels	66;
			Gaps	11;
Qy	32	LAIMTFASLSYSTIV-----DAPLTMSFTTIVVVALLYGRTLLYRRHKFLVPWY	82	
		: : :   :   : :   :   : :   :   : :   :   : :		
Db	9	LALAVANVSTASILVRLAGVGFRAATWRVLSTLTAAILVAAYAVKHKRPVLP----	64	
		: : :   :   : :   :   : :   :   : :   :   : :		
Qy	83	ALLGFVD-----VHGNTLVNKAPELTSTSYSLDCWTIVWSIIFTWMFLG	130	
		: : :   :   : :   :   : :   :   : :   :   : :		
Db	65	--GVDLLVAAAGVALAHFLWMASLPHMSVAMSVTVVDSYPAVLAIVGFRLF-REK	120	
		: : :   :   : :   :   : :   :   : :   :   : :		
Qy	131	YSVYQFVGAAICVGGLL-LVLLSDSGVTA-AGSNPLLGDFLVITGSIILFTSTVGQ	188	
		: : :   :   : :   :   : :   :   : :   :   : :		
Db	121	YTPLQLLGAVIAMGVAGIAFYSSRGASPGGDPKIGFLLAFAFML-----AAVA	175	
		: : :   :   : :   :   : :   :   : :   :   : :		
Qy	189	-----KRDRTEVVMIGVFCGMGISATEITVLERNALSMQWSTGLLAAYVYALSSFL	243	
		: : :   :   : :   :   : :   :   : :   :   : :		
Db	176	KGKLSKYSTLEYTLVPVGFAAIVSLLTTTLVVRAPL-----YG-----YTAETLY	222	
		: : :   :   : :   :   : :   :   : :   :   : :		
Qy	244	CTLTPFLKMSGAAFFNL-----SMLTSDMMAVAIRTFIYNOEVDWLYLAFC	291	
		: : :   :   : :   :   : :   :   : :   :   : :		

```

Db      223  LILLALLPMLGGHTLINLLRRLLSLAANTPIIGEPVGAALAWILLGKVKVSTLEALLMV 282

Qy      292  LVVVGI 297
        :|: ||
Db      283  VVLGI 288

RESULT  12
S58750
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - yeast (Hansenula wingei) mitoc
C:Species: mitochondrion Hansenula wingei
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 07-Dec-1999
C:Accession: S58750
R:Sekito, T.; Okamoto, K.; Kitano, H.; Yoshida, K.
Curr. Genet. 28, 39-53, 1995
A:Title: The complete mitochondrial DNA sequence of Hansenula wingei reveals new char
A:Reference number: S58740; MUID:96022424
A:Accession: S58750
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-567 <SEK>
A:Cross-references: EMBL:D31785
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC3
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2
C:Keywords: mitochondrion; NAD; oxidoreductase

```

Query Match 6.6%; Score 114; DB 2; Length 567;  
Best Local Similarity 20.8%; Pred. NO. 0.084;  
Matches 71; Conservative 59; Mismatches 96; Indels 116; Gaps 16;

Qy 19 LIGLGLQLVSDLAIMTFSASIVSTVDAPLTW-----SFTYTV-----VALLYGTRIL 68  
|| || :|| :|| :|| :|| :|| :|| :|| :|| :|| :||  
Db 149 ILLTSSNNLISFISTELQSYSILYTGIIIPKSKSGSHSLFYVLGGISIIILYGISL 208  
Qy 69 LYRRHKFLPWTYAYALLGFVDHGNYLVNKAPELTSITSVSLDCWTIVWSIITFWPLG 128  
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||  
Db 209 LV-----YITSNIF-INNLINLYSLDIYNNNNILGWLEFI- 243  
Qy 129 TRKSYVFYGAACVCG-----GLLVLSDSGVTAAGS----- 161  
:|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||  
Db 244 -----IIGLKIKGAAPMWNKSILLYSNSNTIITSYLIPKISILSYILLIITLDYN 296  
Qy 162 -NPILGDP-----LVITGSILFTLS----TWQEYCVKRKDRIEVVAM--IGVFGMLI 207  
|| || :|| :|| :|| :|| :|| :|| :|| :|| :|| :||  
Db 297 LNNLNLFNNNNNNNIITYLSIIILISITGSIGGLFOIKIKNLAYSGLNLNGYELLII 356  
Qy 208 SATEITVBERNALSSMOWSTGLAAVVVYALSSEFLTCTLPFLKLMSGAAPFNLSMLTSD 267  
:|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||  
Db 357 ----LSLINNNINS-----ILAVIYY-ITOYCFNHHSIFILLIITAIYNN----- 397  
Qy 268 MWAVAIRFIYQNEVDWLY-----YIAFCL-VVVGIFI 299  
|| || :|| :|| :|| :|| :|| :|| :|| :|| :|| :||  
Db 398 ----NYYNFETINKNLIIYIELNKRYRDLIFCLIIIIIGSFI 435

RESULT 13  
C72378  
Sugar ABC transporter, permease protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 24-Nov-1999  
C:Accession: C72378  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; PMID:99287316  
A:Accession: C72378  
A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-294 <ARN>  
A:CROSS-references: GB:AEO01721; CB:AE000512; MID:g4980922; PID:g4980925; TIGR:TMO419  
A:Experimental source: strain MSB  
C:Genetics:  
A:Gene: TMO419  
C:Superfamily: inner membrane protein ugpA

Query Match      6.5%    Score 112.5; DB 2; Length 294;  
Best Local Similarity 16.1%; Pred. No. 0.055;  
Matches 50; Conservative 65; Mismatches 98; Indels 97; Gaps 14;

QY 48 APLTMSFTTYTVVALLVGTLLLYRRHFLPWPWYWALLGFVDVHGNYLVNKAFE----- 101  
     || : | :: : : : : | | : | : | :  
DB 11 SPLLIFFIVIEVLITVAFTAIDMDYSFETWNVFQNFKDISTDIFIRPIANTFIYT 70  
  
QY 102 -----LTSITSVSILCW-----TIWSIIFTWMELGTKYSV 133  
     | | : : ||| | : : : : ||| | : | :  
DB 71 FGTGLFNLTGALLSLTTISDRGNENFFTLMWLPRLPSPVVYGTLNMFDPTEGL 130  
  
QY 134 YOFVGAAICVGL-----LLVLSDSGVTAAGSNPLLGDFLVTGSILET--LST 181  
     | : : :||| : : : : | | : : : | :  
DB 131 VNFIRGLF---GLPPQDWLHSAPMMWIIFANGFIGAS-----MGM LVPTAAIKS 176  
  
QY 182 VGQEVCVKRKDRIEVMANIGVFGMLISTATEITVELRNALSSMHQSTGLLAAY-VVVALSS 240  
     | : : :||| : : : : | | : : : | :  
DB 177 IPEDY----YYRARID-----GASFPMIVKKITPLIKKHLLLFVTAQTLSLLAS 222  
  
QY 241 FLFLTLPFLFKMSGAAFNLMTSDMWAVAI FTIYNQEVDMLEY-----LAFCLVVV 295  
     | : | : : :||| : : : : | : : : | :  
DB 223 FEY-----ILIITGGPVYR-----TEWWAL-----YTHNAFSHFREGYAALSII LVI I 268  
  
QY 296 GI---FYTK 302  
     | : | |  
DB 269 GVISAFAFMK 278

RESULT 14  
C75131  
hypothetical protein PAB1796 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: C75131  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru  
A:Reference number: A75001  
A:Accession: C75131  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-267 <KAN>  
A:Cross-references: GB:AJ248285; CB:AL096836; MID:g5458067; PIDN:CAB49764.1; PID:e151566  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB1796

	Query Match	6.4%	Score 110.5	DB 2	Length 267
	Best Local Similarity	18.0%	Pred. No. 0.071		
	Matches	55	Conservative	74	Mismatches 122; Indels 55; Gaps 13;
Qy	10	RENEGMRRTLLGLGQLGSFDLAIMTFASLSVSTTVDAPLTMSTFTYTVVVALLYGTILL	69		
Db	3	RKSEGIALALIGMLL-----YGLEPVMVISPNVN-----PISPAF-----FSAFFASLIL	47		
Qy	70	YRRKKEFL-VPWYWT---ALLGFDVGHGKLVNKAFF-----ELTSITSVSILDCWTITWMSIIF	122		
Db	48	IPFADFLEVKAYWRRGALIGF---FGFVLAVLSYFSGAKMSSSVNAAIIPRAEVIPFELL	104		
Qy	123	TWMLFGTKYSVYQFGAAICVGGILLVLLDSGVYTAAGSNPLLGDFLVTIGSILFTLSTV	182		
Db	105	SYAILGGRITRRRTTINSTLVIIIGVILVITQKRTSLN-----VGDFFLLAVPLPFW---QI	156		

Search completed: August 1, 2000, 21:27:51  
Job time: 11219 sec

```

Qy 183 GOEYCVKRRDRIEYVAMI-GVFG--MLISATEITVLERNALSSMOWSTGLLAAYVVVYALS 239
      |      |      |      |      |      |      |      |      |      |      |
Db 157 GHVIAKTNAKPQTIIAFLRNSFGSLFLLLPTLTTELSTFTPSIAEGLIATLGLLIWYSAI 216
      |      |      |      |      |      |      |      |      |      |      |
Qy 240 SFFLFCTLTPTPLKMSGAAPENLSMLTS---DMWAVAIRFIYNQVDWLYYLAFCLVWVG 296
      |      |      |      |      |      |      |      |      |      |      |
Db 217 SRDLSLATAIITPAPAVAIGIALITGVPTLW-----HLLGFFIFITIG 260
      |      |      |      |      |      |      |      |      |      |      |
Qy 297 IFIYTK 302
      |
Db 261 TLNLTR 266

RESULT 15
T01949
hypothetical protein Fl104.14 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01949
R:Abu-Threideh, J.; Stoneking, T.; Langston, Y.; Trevaaskis, E.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of A. thaliana Fl104.
A:Reference number: Z14466
A:Accession: T01949
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-340 <ABU>
A:Cross-references: EMBL:AF096370; NID:g3695372; PID:g3695385
C:Genetics:
A:Map position: 4
A:Introns: 62/2; 101/2; 230/3; 281/2
A:Note: Fl104.14

```

Query Match 6.4%; Score 110.5; DB 2; Length 340;  
Best Local Similarity 22.4%; Pred. No. 0.093;  
Matches 74; Conservative 54; Mismatches 122; Indels 81; Gaps 16;

Qy	32	LAIMTFSASLUSTVTDAPLPMSTTTT	VVVVALLYGTILLYRRHKFLVP----	WYWTALLGF	87		
		! : ! : !	! : ! : !	! : ! : !			
Db	14	LIVSNMIAGMVNAKVKVLDGG	NHVMVIATYRLGISTL-----	FLLPVAYFWERASLMQY	68		
		! : ! : !	! : ! : !	! : ! : !			
Qy	88	VDVHGNYLVNAKELPETSITSV--	SILDCWTIYWSIIFTW--MFLG	TKYSYQVFGNAI-C	142		
		! : ! : !	! : ! : !	! : ! : !			
Db	69	F-----YLLGSLYTSATLGS	AFWAIMPSLTFYVNALIFGFEKLS	TKIGYGVLTLSL	123		
		! : ! : !	! : ! : !	! : ! : !			
Qy	143	VGGILLVL-----LSDSGVT	AAGSNP-----LIGDFLVI	IGSILF-FLSTVG	GEYCVK	189	
		! : ! : !	! : ! : !	! : ! : !			
Db	124	VGGLLLTMYGIIPLTNS	PEQAANSNHTGHE	NWIKGCFLLTGVL	FWSSMMUIQAKINV	183	
		! : ! : !	! : ! : !	! : ! : !			
Qy	190	RKDRIEVAMIGVFGML-----	ISATEITVLERNALSS	MONSTGLLAAY	233		
		! : ! : !	! : ! : !	! : ! : !			
Db	184	YPCPYSTVILSVFGH	LQCALLSLIKTRHLED	WILDELTII-----	TVVIAGV	232	
		! : ! : !	! : ! : !	! : ! : !			
Qy	234	VYVALSFLFC	TL-TPFLKMSG--A	AFPNLSMLTSDM	WAVAIRFTIYNQ	VEDWLYIA	289
		! : ! : !	! : ! : !	! : ! : !	! : ! : !		
Db	233	VAQGM-----CTG	MSWCIAQGPVYSS	SPVPLMS---ATV	FDLILHRIEY	LLGSVIG	284
		! : ! : !	! : ! : !	! : ! : !	! : ! : !		
Qy	290	FCLVVVGIFTY-----	TKTEKDPNNT	310			
		! : ! : !	! : ! : !	! : ! : !			
Db	285	SVVVVIGLYIFLWSR	KOIVECKIMKLP	TNT	315		
		! : ! : !	! : ! : !	! : ! : !			

Search completed: August 1, 2000, 21:27:51  
Job time: 11219 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:27:51 ; Search time 51.03 seconds  
(without alignments)  
567.550 Million cell updates/sec

Title: US-09-147-955-12  
Perfect score: 2475  
Sequence: 1 MVQPHVILTPAQGHINPA.....EGSSHKNLKAFIGDVAKGF 468

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_64:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1207	48.8	467	T02238	glucosyl transferase
2	1090.5	44.1	456	C71402	probable glucosyl
3	1049.5	42.4	458	C71420	hypothetical prote
4	809.5	32.7	479	E71419	probable indole-3-
5	766	30.9	475	F71419	probable indole-3-
6	762	30.8	460	T00639	hypothetical prote
7	761	30.7	484	D71419	probable indole-3-
8	750.5	30.3	471	A54739	indole-3-acetate b
9	727	29.4	456	T00506	indole-3-acetate b
10	712	28.8	438	T00507	indole-3-acetate b
11	645	26.1	453	T00511	indole-3-acetate b
12	553.5	22.4	455	T00584	indole-3-acetate b
13	540	21.8	453	T45603	glucosyltransferas
14	525	21.2	440	T00583	indole-3-acetate b
15	525	21.2	451	T45604	glucosyltransferas
16	522.5	21.1	452	T12981	hypothetical prote
17	500	20.2	462	T01732	UTP-glucose glucos
18	496	20.0	476	T03745	glucosyltransferas
19	491	19.8	476	T03747	glucosyltransferas
20	485.5	19.6	449	T45605	glucosyltransferas
21	477	19.3	466	T07404	probable glucosyl
22	471.5	19.1	490	T46162	glucosyltransferas
23	466.5	18.8	438	T45602	glucosyltransferas
24	444	17.9	447	T12978	hypothetical prote
25	443.5	17.9	488	T07786	UDP-glucose glucos
26	436.5	17.6	478	T08395	UTP-glucose glucos
27	432.5	17.5	487	S41951	UTP-glucose glucos
28	432	17.5	449	S41950	UTP-glucose glucos
29	431	17.4	346	S41954	UTP-glucose glucos

30	427	17.3	478	2	T05423	probable glucosyl
31	423.5	17.1	347	2	T06371	probable UDP-gluc
32	423	17.1	478	2	A71417	hypothetical prote
33	421.5	17.0	420	2	T08005	flavonol 3-O-gluc
34	419.5	16.9	472	2	S39507	glucuronosyl trans
35	417	16.8	471	2	S01052	flavonol 3-O-gluc
36	414.5	16.7	481	2	T01850	UTP-glucose glucos
37	409	16.5	385	2	T12980	hypothetical prote
38	409	16.5	471	2	S01037	flavonol 3-O-gluc
39	408.5	16.5	433	2	S51767	glycosyl transfera
40	407.5	16.5	452	2	G71416	probable glucosyl
41	404	16.3	471	1	S08325	flavonol 3-O-gluc
42	398	16.1	507	2	T46161	glucosyltransferas
43	393.5	15.9	394	2	S41952	UTP-glucose glucos
44	392.5	15.9	287	2	S41953	UTP-glucose glucos
45	391	15.8	452	2	T00981	flavonol 3-O-gluc

## ALIGNMENTS

RESULT 1

T02238  
glucosyl transferase, jasmonate-induced - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 24-Nov-1999  
C:Accession: T02238  
R:Kojima, H.; Hashizume, K.; Imanishi, S.; Nakamura, K.  
submitted to the EMBL Data Library, January 1997  
A:Description: Jasmonate-induced potential glucosyltransferase from tobacco suspensio  
A:Reference number: Z14633  
A:Accession: T02238  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-467 <KOJ>  
A:Cross-references: EMBL:AB000623; NID:d1094897; PID:d1019901  
A:Experimental source: strain BY-2  
C:Genetics:  
A:Gene: JIGT  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 48.8%; Score 1207; DB 2; Length 467;  
Best Local Similarity 50.1%; Pred. No. 1e-82;  
Matches 234; Conservative 87; Mismatches 132; Indels 14; Gaps 7;

Qy 5 HVILTPAQGHINPALQFAKNLVMGIEVTFSTSIYAQSRMDEKSIILNAPK--GLNFI 62

Db 9 HVILALPGQGHINPSLQFSKLLNLGVKVTLSLSSLSAFNR-----IKNLPIEGLTEAP 63

Qy 63 FSDGFDEGFDHSDPV-FYMSQLKCGSETVKIILCSENGQPTICLLYSIFLPWAAEV 121

Db 64 FSDGYDGNFGKGFDDYHLFNSAISKSHGSEFTANLKSAKNGYPFTRVIYITLMDWAGSV 123

Qy 122 AREVHIPALLWSOPATLIDIIYFNHFCYKAMAN-ESNDPNWSITOLPGLPLETRDLP 180

Db 124 AKKHIPSTLFWIOPATVFDIYYRFTNFANFYKNDSDQD--IIEPLSLSSDDP 181

Qy 181 FLLPYGAGKSRLVALPPKELIDTDAETTPKILVNTDELEPEALNAIEGKYFGIGI 240

Db 182 FV--FDDVKSNDAWVESIKRQIETILNSENPRILVNTDELEALNLRVKNVWVGIG 239

Qy 241 IPSAFLGNDPLDASFGGDLFQNSNDYMEWLNSKPNSSVYVIFSGLSMNPISQMEEL 300

Db 240 IPSFSLDEKDRKDRFFAADMIESNNYMEWLDARANKSVIYIAFGSYAEISSQNMEE 299

Qy 301 GLIDIGRFLWVIRENEKKEENKLCIELEKIGIVPWCQVLEKIVKPSLGCFCVSH 360

Db 300 GLLCGRFLWVIRETNGEKPE-EKLTCKDELEKIGIVRWCQVLEKIVKPSLGCFCV 358

Qy 361 CGWNSALESACGVVPAFPQMTDQMTNAQVEDYKVGSRVRINEDGVVSEEEKRCIE 420

Db 359 CGWNSTLESASGVPIVACPIWNDQICNAKLIQDVWKIGVRVNAKKEGIIKRDQKIE 418



Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans  
C.; Chalwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A:Reference number: A71400; MUID:98121113  
A:Accession: F71419  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-479 <BEV>  
A:Cross-references: GB:297339; NID:g22444901; PID:e326931; PID:g22444906  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 32.7%; Score 809.5; DB 2; Length 479;  
Best Local Similarity 38.4%; Pred. No. 5.9e-53;  
Matches 187; Conservative 83; Mismatches 158; Indels 59; Gaps 15;  
QY 5 HVILTFPAOQHINPALQFAKNLVKMGIEVTFSTS-----IYAQSRMDEKSILNAPKGL 58  
DB 8 HVMVSPFGQGHVNPLLRLGLIAASKGLLVTFVTEKPMGKMKRQANKIQGVLR-PVGL 66  
QY 59 NFIP----PSDGFDEGFDHSDKDPVFMYSOLRKCGETVKKIILTCSENGOPITCLLYSIFL 115  
DB 67 GFIREFSDGFADDDEKRFDFAPRPLEAVGQKEINLVK--RYNKEPVTCLINNAFV 124  
QY 116 PMAEVAREVHIPSALLWSQPATILDIYFNHGYEKAMANESNDPNNSIQLPGLPLET 175  
DB 125 PMVCDAVELHIPSALVWQSCACLTAYYYVHRLVKFPT--KTEPDISVEIPCPLLKH 182  
QY 176 RLPSFL---LPYGAKSRLVALPFKELIDTLDAETTPK---ILVNTFDELE----- 222  
DB 183 DEIPSPHPSPYTAFGDI-----ILDQLKFENHKSFYLFIDTFREKIDIMDHM 233  
QY 223 ----PEALNAIEGYFYGIGPLIPSAFLGGNDPLDASFGDLFQNSNDYMEWLNKPNSS 278  
DB 234 SOLCPQAI-----ISPVGPLFKNA-----QTLSSDYKGDISEPASDCMEWLDSPSS 281  
QY 279 VVYISFGSLMPSISQMEIEISKGLIDIGRPLFWIKENEKEENKKGICIEELEKTKG 338  
DB 282 VVYISFGTIANLKQOMEIEAHGVLSGLSVLWVYRPPMEGTFFVPHVLP--RELEERKG 339  
QY 339 IVPWCSQLEVLKHPSLGCVSHCGWNSALESACCPVVPVAPQWTDQNTAKQVEDVWKS 398  
DB 340 IVEWCPQERVLAPAIACFLSHCGWNSMEALTAGVPVPCPQWGDQVTDVYLADVPT 399  
QY 399 GVRV-RINEDGVVESEI--KRCIELVMDGGEGEELRNKAKKWLAREAVKEGSSHK 455  
DB 400 GVLGRGAEEIVSREVVAEKLLEATV--GEKAVELRENARRWKEAEAAVADGSSDM 457  
QY 456 NLKAFID 462  
DB 458 NPKFEVD 464

RESULT 5  
F71419  
probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999  
C:Accession: F71419  
P.; Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dir  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weltzenegger, T.; Pohl, T.M.; Terry, N.; Giel  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans  
C.; Chalwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A:Reference number: A71400; MUID:98121113

A:Accession: F71419  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-475 <BEV>  
A:Cross-references: GB:297339; NID:g22444901; PID:e326932; PID:g22444907  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 30.9%; Score 766; DB 2; Length 475;  
Best Local Similarity 36.9%; Pred. No. 1.1e-49;  
Matches 175; Conservative 91; Mismatches 172; Indels 36; Gaps 14;  
QY 4 PHVILTFPAOQHINPALQFAKNLVKMGIEVTFSTS-----IYAQSRMDEKSILNAPKG 57  
DB 8 PHMVLVSPFGQGHISPLLRGLIATASKGLIVTFVTEEPGLKMKRQANNIQGVLR-PVG 66  
QY 58 LNFIPFSDGFDEGFDHSDKDPVFMYSOLRKCGETVKKIILTCSENGOPITCLLYSIFLPW 117  
DB 67 LGFLRF-EFFEDGFYKEDFDLQKSLVSGKREIKNLVKYK--QPVRCILINNAFVPM 123  
QY 118 AAEVAREVHIPSALLWSQPATILDIYFNHGYEKAMANESNDPNNSIQLPGL-LETR 176  
DB 124 VCDIAEELQIPSALVWQSCACLAAYYYVHQLVKFPT--ETEPEITVDVPKPLTLKHD 181  
QY 177 DLPSLLPYGAGSL-RVALPPFKELIDTLDAETTPKILVNTFDELEALNAIE-----G 231  
DB 182 EIPSPHPSPSSLTGGTILEQIKRLKHPF-----SVLIETTFQLEKDTIDHMSQLCPQ 235  
QY 232 YKFGYIGPLIPSAFLGGNDPLDASFGDLFQNSNDYMEWLNKPNSSVYISFGSLMNP 291  
DB 236 VNFNPIGPLFYMA-----KTRSDIKGDISKPDSCIEWLDSPSSVYISFGTLAFLK 290  
QY 292 ISQMEIEISKGLIDIGRPLFWIKENEKEENKKGICIEELEKTKGIVPWCQLEVLK 351  
DB 291 ONQIDEIAHGLNSGLSLWLRLPPLGLEIAIRPHVLP--ELEEKGIWEMCQOEKVL 348  
QY 352 PSLGCVSHCGWNSALESACCPVVPVAPQWTDQNTAKQVEDVWKSQVVRV--TNEQV 409  
DB 349 PAVACFLSHCGWNSMEALTAGVPVPCPQWGDQVNTAVYIMIDVFTGLRSGASDER 408  
QY 410 VESEI-KRCIELVMDGGEGEELRNKAKKWLAREAVKEGSSHKNLKAFID 462  
DB 409 VPREEVAERLLEATV--GEKAVELRENARRWKEAEASAVAGTSGTSGRNFQFEVD 460

## RESULT 6

T00639  
hypothetical protein F3I6.2 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999  
C:Accession: T00639  
R.; Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Arau  
submitted to the EMBL Data Library, February 1998  
A:Reference number: Z14197  
A:Accession: T00639  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-460 <FED>  
A:Cross-references: EMBL:AC002396; NID:g2749918; PID:g2829862; GSPDB:GN00059; ATSP:F3  
C:Genetics:  
A:Gene: ATSP:F3I6.2  
A:Map position: 1  
A:Introns: 219/3  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 30.8%; Score 762; DB 2; Length 460;  
Best Local Similarity 38.2%; Pred. No. 2e-49;  
Matches 187; Conservative 78; Mismatches 163; Indels 62; Gaps 17;

QY 2 VQPVILTTTPAOGHINPALQFAKNLVKMGIEVTFSTSYIAQSRMBDEKSLILPAKGLNFI 61  
 Db 8 VKGHVILTPYVVOGFAKRLVSKNVKVTIATTTVASSITTPS-----LSVE 60

QY 62 PFSGDG-----EGFDHSDKDPVYNSQLKCCGSETVKKILILTCSENGQPIITCLLYSIFL 115  
 Db 61 PISDGFDFIPIGIPGF--SVDT--YSESFKLNGSETLTLIEKFKSTDSPIDCLLYDSFL 116

QY 116 PWAAEVAREVHIIPSALLWSOPATILDIYYFNHGYEKAMANESN-----DPNWS-IOL 167  
 Db 117 PWGLEVARSMELSAASFNTNLIVCSV-----LRKFSNGDFPLPADPNSAPFRI 165

QY 168 PGLPLLETRDLPSFL-----LPYGAKGSLRVALPFPFKELIDTLDAETPKILVNTFDELEP 223  
 Db 166 RGLPSLYDELPSFVGRHMLTHPEHG--RYLLNQPF-----NHNADWLFWNGFEGLE- 216

QY 224 EALNAIEG---VKFYGIGPLIPSAFLGGNDPLDASFGGDLFQN-SNDYMEWLNSRPNSS 278  
 Db 217 ETQDCENGESDAMKATLIGMIPISAYLDDRMEDDKDYGASLLKPISEKCEHWELETQAOQS 276

QY 279 VVYISFGLMNPSTISQMEETISKGLIDIGRPLFWIKENEGKEENKKGICIELEKIGK 338  
 Db 277 VAFVSFGSFGILEKQAEVAIALQESDLNFWLVIKEAHIAKLPE---GFVESTKDRAL 332

QY 339 IVPWCSOLEVLKHPSLGCFVSHCGWNSALSGLACGVVPAFFPQWTDOMTNAKQVEDVWKS 398  
 Db 333 LVSCNOLEVLAHESIGCFLTHCGWNSLTSLGLVGMVGPQWSQDMNDAKFVEEVWKV 392

QY 399 GVRVRINEDG--VVESEIERKCIELVMDGGEGEELRKNNAKKWKLAREAVKGGSSHKN 456  
 Db 393 GYRAK-EEAGEVIVKSELVRCRLGVME-GESSVKIHESKSKWRDLAVKAMSGSGSDRS 450

QY 457 LKAFIDDVAK 466  
 Db 451 INEFIESLKG 460

RESULT 7  
 D71419  
 probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 A:Variety: columbia  
 C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999  
 C:Accession: D71419  
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk  
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; Giel  
 avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
 Nature 391, 485-488, 1998  
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
 erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans  
 C.; Chaiwatzis, N.  
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
 A:Reference number: A71400; MUID:98121113  
 A:Accession: D71419  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-484 <REV>  
 A:Cross-references: GB:297339; NID:G2244901; PID:e327472; PID:G2244905  
 C:Genetics:  
 A:Map position: 4C09p-4G3845  
 C:Superfamily: flavonol O3-glucosyltransferase

Query Match	30.7%	Score 761;	DB 2;	Length 484;
Best Local Similarity	36.9%;	Pred. No. 2.6e-49;		
Matches 178;	Conservative 87;	Mismatches 178;	Indels 40;	Gaps 15;
Qy	5	HVILTTTPAAGHINPALQFAKNLVKMGIEVTF-STSIYAQSRMDEKSILNA---	PKGLNF	60
	:	:		
Db	13	HVMLVSFGQGVHVNPLRLKGLTASKLLTVFTVTELGWGRMKQANKIVDGLKPVGSGS	72	
	:	:		
Qy	61	IPFSDGDEGF-----DHSKDPVFMVMSQLRRCGSETVKKIILTCSENGQITCLLYSIFLP	116	
	:	:		

Db 73 IRF-EFFDEWAEODDRADPSLYIAHLESVIGIREVSKLVRRYEAEPEVSCLINNPFIP 131

Qy 117 WAAEVARVHIPSALLWSOPATILDIYFENPHGYEKAMAN--ESNDPNWSIQLPGLPLLE 174  
 Db 132 WVCHVAEENIPCAVLWQSCAFSAHYH----YQDGSVSFPPTETELDVKLPVPLK 187  
 Qy 175 TRDLPSEFLPYGAKGSLRVA-LPPFKELIDTDAETTPKILVNTFDELEPALNAIEGY- 232  
 Db 188 NDEIPSEFLHPSRSRTGFRQALIGOLFKNLSKSF-----CVLIDSFDSLEQVEIDYMSSLC 241  
 Qy 233 KFYGTGPIPLPAFAFGGNDPLDASGGDLFQNSNDYMEWLASKPNSSVYVYISFGSLMNPIS 292  
 Db 242 PVKTVGPFUKVA-----RTVTSDVSGDICKSTDKLEWDSRPKSSVYVYISFGTVAYLKQ 296  
 Qy 293 SOMEISKGLDIDIGRPFLLWIKENEGKREENKGLGCEELEKI-----GRIVPMCSOLE 347  
 Db 297 EQIEIANGVLKSGSLFWLVRPPPHDLKVETHVLP--QELKESAKGKGMVDMWCPOEQ 354  
 Qy 348 VLKHPSLGCFVSHCGWNSALESACGVPVPAFPQWTDNNAKQVEDVMKSGVRV--RIN 405  
 Db 355 VLSHPSVACFVTHCGWNSMTESLSSGVVVPCCPQMGDQVTDVAVLIDVFKTVGLRGAT 414  
 Qy 406 EDGVVSESEI-KRCIELVMDGEGEELRKNKAKWELAREAVKEGGSSHKNLKAFIDDV 464  
 Db 415 EERVVPREEVAEKLLEATV--GEKAEELRNALKWAEAAVAPGGSSDKNFRFVEKL 472  
 Qy 465 AKG 467  
 Db 473 GAG 475

RESULT 8  
 A54739  
 indole-3-acetate beta-glucosyltransferase (EC 2.4.1.121) - maize  
 C:Species: Zea mays (maize)  
 C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 29-Oct-1999  
 C:Accession: A54739  
 R:Szerszen, J.B.; Szczeglowski, K.; Bandurski, R.S.  
 Science 265, 1699-1701, 1994  
 A:Title: Iaglu, a gene from Zea mays involved in conjugation of growth hormone indole-3-acetate  
 A:Reference number: A54739; MUID:94367368  
 A:Accession: A54739  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-471 <S2E>  
 A:Cross-references: GB:I34847; NID:g548194; PIDN:AAA59054.1; PID:g548195  
 C:Superfamily: flavonol O3-glucosyltransferase  
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 30.3%; Score 750.5; DB 2; Length 471;  
 Best Local Similarity 35.5%; Pred. No. 1.5e-48;  
 Matches 172; Conservative 85; Mismatches 188; Indels 39; Gaps 13;

Qy 4 PHVILTTPAOGHINPALQFAKNLVKMGIEVTFSTSIYAQSRMD---EKSILNAPKGLNF 60  
 Db 3 PHVLVVPFGGHHMPVQPAKRLASKGVATLVYTRFIQRTADVDAHAPMVEA----- 56

Qy 61 IPFSDGFDE-GFDHSDKDPVFYMSQLKCGSETVKKIILTCSENGQPIITCLLYSIFLPMWA 119  
 Db 57 ---ISDGHDEGGFASAGVAEYLEKQAAASLASLASLVEARASSADAFCTCVYDVSDEVDWL 114

Qy 120 EVAREVHTIPSALLWSOPATILDIYFENPHGY-----EKAMANESNDPNWSIQLP----GLP 171  
 Db 115 PVARWMGLPAPVFPSTQSCAVSAVYHFSGQLRVLAVPPGAAADSGDGGAGAAALSBAFLGLP 174

Qy 172 LLETLDPLSFLPLPYGAKGSLRVALPPFPKILDTDLDAETTPKILVNTFDELEPEALNAIEG 231  
 Db 175 EMERSELPSPFDHGPYPT--IAQAIAKQF---AHAGKDDWVLFNSFEELTEVLAGLTR 229

Qy 232 Y-KFYGIGLPIP-----SAFLGNNDPLDASFGDGLFQNSNDYMEWLNSKPNSSVYVYISFGS 286  
 Db 230 YLKARAIGCPVPLPTAGRTAGANGRI--TYGANLVKPEDACTKWLDTKPDASVAVYISFGS 287

```
Query Match      30.3%; Score 750.5; DB 2; Length 471;  
Best Local Similarity 35.5%; Pred. No. 1.5e-48;  
Matches 172; Conservative 85; Mismatches 188; Indels 39; Gaps 13  
  
y y 4 PHVLTFTTFAQGHNIPALQAKNLVKMGIEVTFTSYAQSRMD---EKSIILNAPKGLNF 60  
|||: || ||||| :|||: | | : |||: |||: |||: |||: |||: |||:  
b 3 PHVLVVFPFCQGHMPWQFARLASKGVATTLVTTFTFIQTADVDAAHPAMVEA----- 56  
  
y 61 IPSDSGDE-GFDHSKDPPFYMSQLRKCGETSVKKIILTCSENGQPITCLLSYIFLPWAA 119  
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
b 57 -ISDGHDEGGFASAAGVAEYLEKQAAAASASILSIVEARSAADAFTCVVYDSVEDWYL 114  
|||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
```

	QY	120	EYAREVHJPSALLWSPATILDIYYFNHG----	EKAMENSDPNNSIQLP-----GLP	171
			:    :	:   :	
	D6	115	PVARMGJLPAVPFSTQCAVSAYVYHFSGRLAVPPGAADGSGAGAAALSEAFGLJP	174	
			:    :	:   :	
	QY	172	LLETRDLSFLPYCAGKSLRVALPPFLELIDTAETTPKTLVNTFEPELEALNATEG	231	
			:     :	:   :     :	
	D6	175	EMERSELFVDHGYPPT--IAMAIIKQF----	AHAGKDDVVLFNSFELETEVLGLTK	229
			:     :	:   :     :	
	QY	232	Y-KFYGIQPLIP-----SAFLGNPDLDASFGGDLFONSNNDYMELWNKSPNSVVVISFGS	286	
			:     :	:   :     :	
	D6	230	YLKARAIGCVPLPTAGRATAGANGRI--TYGANLVKPEDACTFKWLDTKPDRSVAIVSFGS	287	
			:     :	:   :     :	



A;Note: T27E13.12  
C;Superfamily: flavonol O3-glucosyltransferase

Query Match 22.48; Score 553.5; DB 2; Length 455;  
Best Local Similarity 32.29; Pred. No. 8e-34;  
Matches 154; Conservative 79; Mismatches 182; Indels 65; Gaps 16;

Qy 5 HVILTTPTAOGHINPALQFAKNLVKM--GIETVETS-----IYAQRMDKSI LNAPK 56  
Db : : ||||| : : || : : | : : | : : | : : |  
Qy 13 HVAAMPYGRGHINPMNKLCKRLVRRPNLHVTFVTEWLGFGDPKPDRIHFTLP- 71  
Db : : ||||| : : || : : | : : | : : | : : |  
Qy 57 GLNFIPSDGDFEDGHSKDVPFYMSQLRCGSTVKKIILTCSENGOPIICLLYSIFLP 116  
Db : : || : : | : : | : : | : : | : : | : : | : : |  
Qy 72 --NLIP-----SELVRAKDFIGFDIVYTRLEPPEKLL--DSLNSPPPSVIFADTVVI 121  
Db : : || : : | : : | : : | : : | : : | : : | : : |  
Qy 117 WAAEVAREVHIPSSALLWSOPATILDIYFN---FHGYEKAMANESNDPNWSIQPLGLPL 172  
Db : : || : : | : : | : : | : : | : : | : : | : : |  
Qy 122 WAVVRGRKRNPVVSMTSATILSFSLHSDLLISHHALFEPSEEVEVDY---VPLGSP 178  
Db : : || : : | : : | : : | : : | : : | : : | : : |  
Qy 173 LETRDLFSFLPYCAKGASLRVALPFPEKLDLTDAETTPKILVNTFEDEALNALEG- 231  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 179 TKLRDLPIDFGYSDR-VFKTAKLCFDEL-----PGARSLFFTAYELEHAIDAITSK 231  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 232 --YFYGIGPLIPSAFTGGNDPLDASFDDLFQNSN---DYMEWLNSKPNSSVVYISFGS 286  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 232 LDIPVYAGPLIPPEELS-----VQDNKEPNYIQMLESGPEGSVLYISQS 278  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 287 LMNPSIQMESISKGLIDIGRPFTLVWKENKGEENKKLGCIIEEL-KTGKIVPWCSQ 345  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 279 FLSVSEAQMEEIVKGLRESGVRFLLWARGGE-----LKLKAELGSLGVVVSWCDDQ 329  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 346 LEVLKHPSLGCFCVSHCGWNSALESACGVPVAFQWTDONTNAKOVEDVWKSQVRYRIN 405  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 330 LRVLCHKAVGFHTCCGNSTLEGIIYGVPMPLRPLWDQILNAKNIVEDWRVGMRERT 389  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 406 EDG--VVVESEIEIKRCIELVMD-GGEKGEEELRKNAKWKELAAREAVKGGSHKLNKAFI 461  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 390 KKNELLIGREIKEVVRHFMDRESEEGKEMRRACDLSEISRGA VAKSGSNVIDEFV 448  
Db : : ||||| : : || : : | : : | : : | : : | : : |

RESULT 13  
T45603  
glucosyltransferase-like protein - Arabidopsis thaliana  
N:Alternate names: protein F12A12.180  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000  
C:Accession: T45603  
R:Choiene, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; submitted to the Protein Sequence Database, December 1999  
A:Reference number: Z23008  
A:Accession: T45603  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-453 <CHO>  
A:Cross-references: EMBL:AL13314  
A:Experimental source: cultivar Columbia; BAC clone F12A12  
C:Genetics:  
A:Map position: 3  
A:Introns: 161/1  
A:Note: F12A12.180  
C;Superfamily: flavonol O3-glucosyltransferase

Query Match 21.88; Score 540; DB 2; Length 453;  
Best Local Similarity 31.48; Pred. No. 8.1e-33;  
Matches 153; Conservative 83; Mismatches 171; Indels 80; Gaps 19;

Qy 6 VILTPTAOGHINPALQFAKNLVKMGIEVT-FSTSIYAQSRMDEKSI LNAPKLNFIIPS 64  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 10 VLVPFPQAQGHISPMQMLAKTLHLKGSFITVVFQKFNYFSPSDDFT-----HDFQFVIP 64  
Db : : ||||| : : || : : | : : | : : | : : | : : |  
Qy 65 DGFDGEDPHSKDVPFYMSQLRK-CG---SETVKKIILTCSENGOPIICLLYSIFELPWAAE 120  
Db : : ||||| : : || : : | : : | : : | : : | : : |



```

Db      160  PGLSPTRLSDLQ---LLNGYSHQVNFNFKKSGELY-----KAKYLLFPSAVELEPKAID 211
Qy      228  AIEG---YKFGYIGLIPSLAFLG-GNDPLDASFGGLFQNSNDYMEWLMSKPNSSVYIIS 283
Db      212  FFTSKDFPVYSTGLPILEELSGVNE-----NRELDYFKWLDEQPSSVLYIIS 260
Qy      284  FGSLLMNPISISOMEIISKGLIDIGRPFLWVIKENEKQKEENKKLGCIIELE-KIGKIVPW 342
Db      261  QGSFLSYSEAQMEETIVGVREAGVKFFPWARGCE-----LKLEALEGSLGVVWSW 311
Qy      343  CSQLEVLKHPISLGCFSVSHCWNSALESIACGVPVAPFQWTDQMTNAKQVEDVWKSGVRV 402
Db      312  CDQLRVLCCHAAIGGFVTHCGYNSTLEGICSGVPLLTFFPMDQDFLAKAKIVESWRVGMGI 371
Qy      403  RINE-DGVVSESEIKRCIELVMDG-GEKGEELRNKNAKKWKEALARAEVGGSSHNKLA 459
Db      372  ERKQOMELLIVSDEIKELVKRPMDDGESEEGKEMRRRTCDLSEICRGAVAKGGSSDANIDA 431
Qy      460  FIDDVAK 466
Db      432  FIKDITK 438

RESULT 15
T45604
glucosyltransferase-like protein - Arabidopsis thaliana
N:Alternate names: protein F12A12.190
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Feb-2000
C:Accession: T45604

```

Submitted to the Protein Sequence Database, December 1999

A: Reference number: T23008  
A: Accession: T45604  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-451 <CHG>  
A: Cross-references: EMBL:AL13314  
A: Experimental source: cultivar Columbia; BAC clone F12A12  
C: Genetics:  
A: Map position: 3  
A: Introns: 160/1  
A: Note: F12A12.190  
C: Superfamily: flavonol O3-glucosyltransferase

	Query Match	21.2%	Score 525;	DB 2;	Length 451;	
	Best Local Similarity	30.8%	Pred. No. 1-le-31;			
	Matches 146;	Conservative	84;	Mismatches 198;	Indels 56;	Gaps 14;

Qy    6   VILTTFPAGGHINPALQFAKNLVKMGIEVTFSTSYIAQRMDKSIILNAPKGLNPIPSD 65  
     | : ||||| : | : | : | : | : | : | : | : | : | : | : | : | :  
Db   10   VYLVAVPAGGHISPTMLAKTILHLKGFSITI-----AQTKFNYPSPSDDFDFOFVTIPE 64

[illegible]

QY 348 VLKHPSLGCFVSHCGWNSALESLACGVPVYAFPPQWTDQMTNAKQVEDVWVWKSQVVRINED 407  
Db 337 VLHPAVGGFWSHCGWNSTLESIGEGVPMICKPFPSSDQWVNRYLECVWVWKSQVVRINED 392  
QY 408 GVVESEIIRKCIELVMDGGEKGEELRKNNAKWKELAREAVKEGGSSHKNLKAFI 461  
Db 393 GDLDRGAVERAVRRLN-VEEEGEGMRKRAISLKEQLRASVISGGSSHNSLEEFV 445

Search completed: August 1, 2000, 21:27:53  
Job time: 11221 sec

[illegible]

```
QY 301 aaacatacaagtttttaatttttccactaagcaagaaaaatatggtgcagcctcatgtcat 360
DB 301 AAACATACAGTTTTTAATTTTCTACTAAGCANGAAAAATATGGTCAGCCTCATGTCTAT 360
QY 361 cttacaaacatttccagcaagggccattatccagcacttcaatttgcgaagaatct 420
DB 361 CTTAACACATTTCCAGCACAGGCCATATTAATCCAGCACTTCAATTTGCCAAGAATCT 420
QY 421 tgtcaagatgggcataagaagtgacattttctacaagcatttatgcccgaagccgtatgga 480
DB 421 TGTCAAGATGGGCATAGAAGTGACATTTTCTACAAACATTTATGCCCAAGCCGTATGGA 480
QY 481 tgaataatccattcttaatgcacaaaggattgaatttccatttccatttccgatggctt 540
DB 481 TGAATAATCCATCTTAATGACCAAAAAGGATGAATTTCAATTCATTTCCGATGGCTT 540
QY 541 tgaatgaagttttgatcatctcaaaagaccctgtattttacatgtcacaacttgcataatg 600
DB 541 TGATGAAGTTTGTGATCATTCAAAAGACCCTGTATTTTACATGTCCAACTTCGTAATG 600
QY 601 tggaaatgaactgtcaaaaaataattctcaacttgcctctctgaaaaatggacagcctataac 660
DB 601 TGGAAATGAACGTGTCAAAAAATAATTCTCACTTCTCTGAAAAATGGACAGCCTATAAC 660
QY 661 ttgcctcaactttactccatttccctctcttgggcagcagaggttagcagtgaaagttcacat 720
DB 661 TTGCCTCACTTTACTCCATTTTCTTCTCTTGGCGCAGCAGAGGTAGCAGTGAAGTTCAAT 720
QY 721 ccctctgctctcttcttggagtcacaccagcaacaatattggacatatattacttcaactt 780
DB 721 CCTTCTGCTCTCTTTGGAGTCAACCAGCAACAATATTGGACATATATTACTTCAACTT 780
QY 781 tcatgatataaagactatggctaataatgaatccaatgatccaattgttccattcaact 840
DB 781 TCATGATATGA AAAAGCTATGGCTAATGAATCCAATGATCCAATTTGGTCCATTCAACT 840
QY 841 tcccgggtctccactactgaaactcgagatctctctcattttacttctcttatggtgc 900
DB 841 TCCCGGGCTTCCACTACTGAAACTCGAGATCTTCTCTCAATTTTACTTCTTATGGTGC 900
QY 901 aaagggagcttcctgagtgacacttccaccattcaagaattgatagacacattagatgc 960
DB 901 AAAAGGAGCTTCTCGAGTTCGACTTCCAGCATTCAAGAATTTGATAGACACATTAGATGC 960
QY 961 tgaaccactcctaagactctgtgaatcatttgatgaattagaccctgaggaactcaa 1020
DB 961 TGAACCACTCTCTTAAGATTCTTGTGAATACATTTGATGAATTAGAGCCTGAGGCCTCAA 1020
QY 1021 tgcattgaaggttataagttttatggaattggaccgttgattccctctgcttcttggg 1080
DB 1021 TGCATTTGAAGTTTATAGTTTATGGAATTGGACGTTGATTCCTTCTGCTTCTTGGG 1080
QY 1081 tggaaatgacctttagatgcttcatttgggtgctatcttttccaaattccaaatgacta 1140
DB 1081 TGGAAATGACCCCTTATAGTCTTCAATTTGGTGTGATCTTTTCAAAATTCAAATGACTA 1140
QY 1141 tatggaatggttaaactcaaaagccaaattcatcagttgttttatatatcttttgggagctt 1200
DB 1141 TATGGAATGGTTAAATCTAAAGCCAAATTCATCAGTTGTTTATATATCTTTTGGGAGTCT 1200
QY 1201 aatgaatccatctattagcacaatggagagatatcaaaagggttgatagacataggaag 1260
DB 1201 AATGAATCCATCTATTAGCCAAATGAGGAGATATCAAAAGGGTTGATAGACATAGGAAG 1260
QY 1261 gccgtttttatgggtgataaaagaaatgaaaaaggcaagaagaagaataaaaaagct 1320
DB 1261 GCCGTTTTTATGGGTGATAAAAGAAATGAAAAAGGCAAGAGAGAGAAATAAAAAGCT 1320
QY 1321 tgggttatgaagaattggaaaaaatagaaaaaatagttcccatggtgttcacaacttga 1380
DB 1321 TGGTGTATTAAGAATTTGAAAAAATAGAAAAAATAGTTCCTCATGTTTCAACAACTTGA 1380
```

```
QY 1381 agttctaataacatccatcttaggatgttttcttctcattgtggatgggaattcagcctt 1440
DB 1381 AGTTCCTAAACATCCATCTTTAGGATGTTTTTGTTCATTTCTCATTTGGATGGAATTCAGCCTT 1440
QY 1441 agagagtttagcttgtggagtgccagtttggcatttccctcaatggacagatcaaatgac 1500
DB 1441 AGAGAGTTTAGCTTGTGGAGTGCCAGTTGTGGCATTTCCCTCAATGGACAGATCAAAATGAC 1500
QY 1501 aaatgccaaaacaagttagaagatgttggaaaaatggagtaagaagtgaagaataaatgaaga 1560
DB 1501 AAATGCCAAAACAAGTTGAAGATGTGGAAAAGTGGAGTAAGAGTGAAGATAAATGAAGA 1560
QY 1561 tgggtgttgaagtggagaaatcaaaagtgatttgaattgtgaatgtgaatgtgagagaga 1620
DB 1561 TGGTGTGTGAAAGTGAGGAAATCAAAAGTGCTATTGAATTTGGTAATGGATGGAGAGA 1620
QY 1621 gaaaggggaagaattgagaagaatgctaaagaatggaaatggaaatggctgagaagctgt 1680
DB 1621 GAAAGGGGAAGAATTGAGAAAGAAATGCTAAGAAATGGAAGAATTTGGCTAGAGAAGCTGT 1680
QY 1681 gaaggaagtggtcttccacacaagaatttaaaagcttttatttatttatttgcataatgttcccaagg 1740
DB 1681 GAAGGAAGTGGAATCTTCACACAAGAAATTTAAAGGCTTTTATTGATGATGTTGCCAAAGG 1740
QY 1741 gtttataatttaccaggttttccgtgatatctacttccctctagttggcgattcactctt 1800
DB 1741 GTTTTAATATTACAGGCTTTTGGCGTGATATTACTTTCCCTAGTTGGCGCATTCACCTTT 1800
QY 1801 tgtggaacttcttgacaaaaaactgagggaatgtgctaaagacacogctaaatgctttaagaa 1860
DB 1801 TGTGGACTTCTTTCACAAAAAACTGAGGGAATGTGCTAAGACACACGCTAATGCTTTAAGAA 1860
QY 1861 gtcatttccaaagcttgaagcctctttttaaactatttagccagtaataatatagggttc 1920
DB 1861 GTCATTTCCAAAGCTTGAAGCCTGCTTTTAAACTTATTAGCCAGTAATCTATAGGGTTC 1920
QY 1921 tctctatttttctgtctctctctttttagccttttcttccaaagttttaaagaatagcg 1980
DB 1921 TCTTCTATTTTCTCTGCTCTCTTTTAGCCCTTTTCTTTCCAAAGTTTAAAGAAATAGCG 1980
QY 1981 tgaacatagcttagttagtcttctgtatctctctctctctctctcttaccagtgcaagattatgctt 2040
DB 1981 TGAACATAGCTTAGTAGTAGTCTTGGTATCTCTATCTTACCAAGTGCAAGATTATGCTT 2040
QY 2041 atcctctccctcaatttcttaataaaatgcaagaatgaaaaagtcacaaaaaagaaaaa 2100
DB 2041 ATCTGTCTCTCCCTAAATTTCTTAAATGAAGATGAAGAAAGTACAAAAAAGAAAAA 2100
QY 2101 aaaaa 2105
DB 2101 AAAAA 2105
RESULT 2
X02826
ID X02826 standard; cDNA; 1506 BP.
AC X02826;
DF 14-MAY-1999 (first entry)
DE W0905287 Seq ID 1.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Perilla frutescens.
FH Key Location/Qualifiers
FT CDS 17..1399
FT /tag= a
FT /product= "protein with flavonoid 5-transglycosylation activity"
```

PI W0905287-Al.  
PN 04-FEB-1999.  
PF 16-JUL-1998; J03199.  
PR 25-JUL-1997; JP-200571.  
PA (SUNR ) SUNTORY LTD.  
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y, Yamazaki M;



Query Match 20.9%; Score 439.8; DB 1; Length 1474;  
Best Local Similarity 61.0%; Pred. No. 5.4e-87;  
Matches 865; Conservative 0; Mismatches 497; Indels 57; Gaps 7;

QY 337 aaatagtgtgagccatctgtcattccttaacacacattccagacacagggccattatcaac 396  
DB 25 AAAAAATGGTCCGCGCGGTGCTGTAGCAACGTTTCCGCGCAAGGCCACATAATCC 84

QY 397 agcaactcatttccaagaactcttcaagatgggcatagaaagtacattttctcaag 456  
DB 85 CGCCCTCCAAATCGCCCAAGAGACTCTAAAGCCGCACGTACGCTCACGTTTTCACGAG 144

QY 457 cattttggccaaagccgatgatgaataaaatccattcttaagc-----accaaaag 510  
DB 145 CGTTTATGCATGCGCGCGGATGCCAACACAGCCTCCGCGCTGCCGGAACCCACCGGG 204

QY 511 attgaatttcattccatttccgatggctttgatgaaggtttttgatcattcaaaagacc 570  
DB 205 CCTCGACTTTCGTGGCTTCTCCAGCGCTACGACGAGCTGAAGCCGGCTGAAGCCCGCGCGCAGG 264

QY 571 tgtattttacatgtcacaacttcgttaaatgtggaagtgaactgtcaaaaaataattct 630  
DB 265 GAAGCCCTACATGTCGAGATGAAGCCCGCGCTCCGAGCGCTTAAGAAACCTCCTTCT 324

QY 631 cacttgcctgaaatggagacgctataacttgcctacttacttactcatttcttctctg 690  
DB 325 CAACACGAGCA-----CCTCACTTTCGTGCTACTCCCACTCTTTGCATG 372

QY 691 ggcagcagaggtagcagtgaaagtccacatcccttcttcttcttggagtcacaacagc 750  
DB 373 GCGCGGAGGTGGCGGTTGTCCACAGTCCGACCGCCCTTCTCTGGTCCGAGCCCGC 432

QY 751 aacaaatattggacatatattacttcaacttcaatgatgatgaataagctatggctaatga 810  
DB 433 CACCGTGTGTGATATACCACTTCTACTTCAACGGCTACGCAGAGAGATCGACCCCG 492

QY 811 atccaatgatccaaatgttgccattccacactccggctccactacttggaaactcgaga 870  
DB 493 TTCATATGAA-----ATTCAGCTCCCTCGCTTCCATCTCCGAGCAGCGCAG 540

QY 871 tcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 930  
DB 541 TCTTCCGAGCTTCTCTGCTGCC-TGCAGCGCGGAGAGATTCCCGTTGTATGA----- 590

QY 931 attcaagaattgatagacacattagatcctaaacccactcctcaagattcttgatc 990  
DB 591 --TGAAGGAGAACTGGAACTTTAGACGGTGAAGAGAGAGCCGAAAGTATTGTTGAACAC 648

QY 991 atttgatgaattagagcctgagcactcaatgcaattgaaggttataagttttatggaat 1050  
DB 649 GTTTGATGCTGGAGCGCGATGCATCAGCGCTATTGATGATGATGATGATGATGATGAT 708

QY 1051 tggacgcttgattccttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1110  
DB 709 CGGCGGTTGATTCCTCCGCTTCTTGGACGCGGAAGATCCCTCCGAAACGCTTACGG 768

QY 1111 tgggtatcttttcaaaatcaaatgactata-----tgggaatggtttaaactcaagcc 1164  
DB 769 CGCGCATCTTTTCGAAATAATCGAGGAGAAATACTCCGTGGAGTGGTTGAACCTCGAAGCC 828

QY 1165 aaattcatcagttgtttatatcttcttggagtcataatgaatccattctattagccaat 1224  
DB 829 GAATCTTCGGTGGTGTATGTCTGCTTTGGACCGTTTGGATTTCCAAAGCCACAAAT 898

QY 1225 ggaggagatatcaaaaggggttagatagacataggaagccggttttttgggtgataaaga 1284  
DB 889 GGAAGAGATTGGAAAGGGCTATTAGCTTCGCGAAGCCCTTTTATGATGATACGAGA 948

QY 1285 aaatgaaaagcc-----aaagaagaagaataaaagcttgggttattgaaga 1335  
DB 949 ACAGAGAATACGACGCGCAAGAGAAAGAAAGAAAGAGAGTGTAGTTGCATTGGGA 1008

QY 1336 attggaaaaaataggaaaaatagtttccatgtgttcacaaactgaagtttcaaaaacatcc 1395  
DB 1009 ATTGAAAAAATGGGAAAAATAGTGTCTGTTGCTCGCAGTTGGAGGTTCTGGCGCACCC 1068

QY 1396 atcttttagatgttttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1455  
DB 1069 TGGGTTGGGATGTTTCTGTCAGCATTTGTGGTGGAACTCGGCTGTGGAGAGCTTGAGTTG 1128

QY 1456 tggagtgcaggttggcatttctcctcaatggacagatcaaatgacaaatgacaaatga 1515  
DB 1129 CGGGATTCGGGTGGCGGTGCCGAGTGGTTGTATGACACCAAGCAATCGAAGCTGAT 1188

QY 1516 tgaagatgtgtgaaagtggatgaagtgagaaataaaatgaagatgggtgttcttctgaag 1575  
DB 1189 TGAGGATCGTGGGGACAGGGGTGAGAGTGAGAAATGAAGGGGTGGGCTTGATGG 1248

QY 1576 tgaggaaatcaaaaggtgtattgaattggtaattggatggaggagagaaaggggaat 1635  
DB 1249 ATGTGAGATAGAAAGGTGTGTGGAGATGGTATGGATGGGGTGAACAAGACCAACTAGT 1308

QY 1636 gagaagaatgctaaagaatggaagaatggctagagaagctgtgaaggaaggtggatc 1695  
DB 1309 GAGAGAAATGCCATCAATGGAAGACTTTGGCCAGACAAGCCATGGATAGGATGATC 1368

QY 1696 ttcacacaagaatttaaagccttttattgatgttgc 1734  
DB 1369 TTCCTCAACAATCTCAACGCCCTTCTTCGTCAAGTTGC 1407

RESULT 4  
X02828  
ID X02828 standard; cDNA; 2062 BP.  
AC X02828;  
DT 14-MAY-1999 (first entry)  
DE W0905287 Seq ID 3.  
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; da.  
OS Verbena hybrida.  
FH Key Location/Qualifiers  
FT CDS 26..1411  
FT /tag= a  
FT /product= "protein with flavonoid 5-transglycosylation activity"

W0905287-A1.  
W0905287-A1.  
PD 04-FEB-1999.  
PF 16-JUL-1998; J03199.  
PR 25-JUL-1997; JP-200571.  
PA (SUNR ) SUNPORT LTD.  
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y, Yamazaki M;  
PI WPI; 99-142940/12.  
DR P-PSDB: W92949.  
DR Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration  
PT Disclosure; Page 56-60; 89pp; Japanese.  
PS This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut flowers harvested from them.  
CC Sequence 2062 BP; 608 A; 354 C; 508 G; 592 T;

Query Match 20.2%; Score 424.2; DB 1; Length 2062;  
Best Local Similarity 60.2%; Pred. No. 1.4e-83;  
Matches 856; Conservative 0; Mismatches 513; Indels 54; Gaps 7;

QY 329 aagcaagaataatggtgcagcctcatgtctatcttcaacacatttccagacagagccat 388  
DB 14 AATAAAAAAATGACGAGCTACGTCCTCTTGGCCACATTCACGACAGGACAC 73

QY 389 attaatccagcacttcaatttggcaagaatcttctgaagatgggcatagaagtacatt 448  
DB 74 AATAATCCCGCCCTTCAATTCGCAAGCGGTCTCGCAAAATGCCGACATTCACATTC 133

```
Qy 449 tctacaagcatttatgcccaaaagccgatgatgataaaatccattcttaatgcacaaaaa 508
Db 134 TTCACCGCGTCTAGCATGCGCGCGCATG--TCCAGAACCGCGCTGGCTCAACCGG 190
Qy 509 gnatgaattccatccatcttcogattgcttggatgaagtttttgatcatcctcaaaagac 568
Db 191 CTCATCAATTTTGTCTGCTTTTCCGACGGGTATGACGAGGTTACAGCCGGAGACGAT 250
Qy 569 cctgtattttacatgctcaacttcgtataatgtggaagtgaactgtcaaaaaataatt 628
Db 251 GGAAGAAGACTACATGTCGGAGTGAAGACGACAGGTATAAAAGCCTTGAGCGATACTCTT 310
Qy 629 -----ctcaactgtctgaaattgacagcctataaacttgcctacttacttcaacttttc 682
Db 311 GCAGCCATATGTCGATCAAAAAGACGACAAATCAGCTTCGTGGTGTACTCCACCTC 370
Qy 683 ctctcctgggcagcagaggtgacgctgaagttccactccctctctctctcttttggagt 742
Db 371 TTTGCATGGCGGCCAAGGTGGCGGTGAGTTCCATCTCCGAGGCGCTACTCTGGATT 430
Qy 743 caaccagcaacaattatggacatatattacttcaacttccatcttcgatatagaataagctatg 802
Db 431 GAGCAGCTACGGTGTGGATATATTTTACTTTTATTTTCAACGGGTATAGCGCAAGAAATC 490
Qy 803 gctaagtgaatccaatgacccaattggccattccaactcccgggctccactactcgaa 862
Db 491 GATGGGGTTCGGATGCTATTACATTGCCC-----GGAGGACTCCAGTGTGCGCC 541
Qy 863 actogagatcttccctctatttacttccctctatggtgcataaaggagttctcagttgca 922
Db 542 CAGCGTGATTTACCGTCTTCTCTCTCTCTT-----CCACGATGAGAGATTC 589
Qy 923 ctccacacattcaaaagaattgatagacacattagatgctgaacacactccctaagatttt 982
Db 590 CTTTCACTGATGAAGGAGAAATTTGAAACTTTAGAAAGGTGAAGAAAAACCTTAAGTCTGTG 649
Qy 983 gtgaatacatattgatgaattagagcctgagcactcaatgcaatgaagggttataagttt 1042
Db 650 GTGACACCTTTGATGCGTGGAGCCTGATGGCTCAGGGCATTGATGAATACGAGATG 709
Qy 1043 tatggaattgaccgttgattccctctctctcttcttgggtggaatgaaccttttagtct 1102
Db 710 ATTGCAATCGGCGCGTGTATCTCTCCGATCTTTGGAGCGGTAAAGATCTCTCGACAGG 769
Qy 1103 teatttgggtgatcttttttcaa-----aattcaaatgaactatataaggaatggta 1153
Db 770 TCTTTCCGCGGAGATTTGTTGAGAAAAGGGTGCNAATGACGAGATTCGCTCGAATGGTTG 829
Qy 1154 aactcaaaagccaaatctcatcagttgttttatatatcttttggagtgctaatgaatccatct 1213
Db 830 AGCAGCAATCTCGATCTTCGGTGTGTTTACGTTTCGTTCCGAGGCTTGTAAATACGACG 889
Qy 1214 attagccaaatgagagatatcaaaagggttgatagacataggagccgctttttatgg 1273
Db 890 AAGTCGCAAAATGGAAGATAGAGAGGGCTGTTAGATTGTTGGAGGCGGCTTTTGTGG 949
Qy 1274 gtgataaagaagaaatgaaaaagcagaagaagaataaaaaagcttggttgtattgaa 1333
Db 950 GTGGTAAGAGTA-----AACGAAGGAGAGAGGATATTGATAAGTTGCGATGGAG 997
Qy 1334 gaattggaaaaataaggaaaaatgctccatggtgttcacaaactgaaagttctaaataat 1393
Db 998 GAGTTGAACAGACTGGGNAATTTGTATCTTGTGTTCTCAATTTGGAAGTCTCGACGAT 1057
Qy 1394 ccatcttttagtagttttgttctctcatgtgagatgaattcagccttagagagtttagct 1453
Db 1058 CCTCTGTTGGGATGTTTCGTGACACACTGCGGGTGGAAATTCGACTCTAGAGAGTATATCT 1117
Qy 1454 tgtgagtgcaagtgtgtgaccttctcactcaatggaacagatcaaatgacaaaatccaaaca 1513
Db 1118 TTCGGGGTTCGAGTGGTCTTTTCCGAGTGGTTCCGATCGATCAAGGGAGCAATTCGAAGCTG 1177
```

```
Qy 1514 gttgaagatgtgtgaaaaagtgagtgagtgagaataaaatgaagatggt---gttgtt 1570
Db 1178 ATGGAGGATGTGTGAGGACGGTGTGAGAGTGTAGAGCTAATGAGAGGGTAGCGTCGTT 1237
Qy 1571 gaaagtagggaaaaatacaaaagtgattgaattggttaattggtgagggagagaaaaagggaa 1630
Db 1238 GATGTTGATGAAATTTAGGAGATGTATTGAGGAGGTTATGGATGGGGGAGAAAAGAGTAGG 1297
Qy 1631 gaattgaaaaagaatgctaaagaatggaagaattggctagagaaagctgtaagaagaagt 1690
Db 1298 AAATTAGAGAGTGTCTGGCAAGTGGAGGATTTGGCAAGAAAAGCTATGGAGGAAGAT 1357
Qy 1691 ggaattccacacaagaatttaaaggctttttattgatgatgttg 1733
Db 1358 GGATCTTCAGTTAACACCTCAAGTCTTCTTGTGATGAGGTTG 1400

RESULT 5
X02829
ID X02829 standard; cDNA; 1671 BP.
AC X02829;
DE 14-MAY-1999 (first entry)
PF WO9905287 Seq ID 4.
KW Plant; flavonoid 5-transglycosylation activity; 5Tg; variety; colour; ds.
OS Torenia hybrira.
FH Key Location/Qualifiers
FT CDS 45..1481
FT /*tag= a
FT /product= "protein with flavonoid 5-transglycosylation
activity"
WO9905287-A1.
PD 04-FEB-1999.
PR 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR ) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI; 99-142940/12.
DR P-PSDB; W92950.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PT used to transform plants for improvement of plant coloration
PS Disclosure; Page 60-64; 8pp; Japanese.
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylation (5Tg) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC the production of flower varieties with desired coloration and cut
CC flowers harvested from them.
SQ Sequence 1671 BP; 530 A; 344 C; 399 G; 396 T;

Query Match 13.7%; Score 289; DB 1; Length 1671;
Best Local Similarity 56.1%; Pred. No. 3.4e-54;
Matches 847; Conservative 2; Mismatches 557; Indels 105; Gaps 12;

Qy 329 aagcaagaaaaatggtgagcctcatgctcttaacaacatttccagcacaagccat 388
Db 36 AAAAAAAAATGGTTAAACAACGCCCATATTCTACTAGCAACATTCACACACACAGGCCAC 95
Qy 389 attaatccagcacttcaatttcccaagaatcttctgaagatgggcatag-----aagtg 442
Db 96 AFAAACCCCTTCTCTCGAGTTCGCCAAAAGGCTCCCTCAACACGGGATACGTCGACCAAGTC 155
Qy 443 acattttctacaagcatttatgccccaaagccgatgatgataaataatccatttcaatgca 502
Db 156 ACATTCTTCACGAGGTATACCATTTGAGACGCATGCGCTTCGAACCCGATCCGAGC--- 212
Qy 503 ccaaaagagtgaaattcaattccatttccgatggcttttgatgaaggttttgatcattca 562
Db 213 ---ACAGAAATCGATTTCTCGCAATKTYCAGATTCTTACGATGATGGCTTAAGAAGGC 269
Qy 563 aaagaccctgtattttacatgtcacaacttcgtaaatgtggaagtgaactgtcaaaaaa 622
Db 270 GACGATGGCAAAACTACATGTGCGGAGTGAAGAAAGCGCGGAACGAGCCCTTAAGGAC 329
```





```
RESULT 7
ID V02042 standard; DNA; 7287 BP.
AC V02042;
DE 08-JUN-1998 (first entry)
DT Plasmid pWRG3169 encoding murine interleukin-12.
KW Interleukin-12; IL-12; cytokine; growth factor; mouse;
KW Plasmid pWRG3169; cancer; tumour; metastasis; gene therapy; ds;
KW cyclic; circular.
OS Chimeric - Mus musculus.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Bos taurus.
OS Chimeric - Rhesus macaque polyoma virus.
FH Key Location/Qualifiers
FT Promoter
FT 1..628
FT /*tag= a
FT /note= "CMV promoter"
FT IDNA 629..810
FT /*tag= b
FT CDS 953..1673
FT /*tag= c
FT /product= p35 subunit
FT /note= "contains an intron"
FT Intron 1259..1331
FT /*tag= d
FT polyA_site 1797..2024
FT /*tag= e
FT /note= "bovine growth hormone polyA site"
FT 2110..2737
FT /*tag= f
FT /note= "CMV promoter"
FT IDNA 2738..2919
FT /*tag= g
FT CDS 2983..3990
FT /*tag= h
FT /product= p40 subunit
FT 4075..4306
FT /*tag= i
FT /note= "bovine growth hormone polyA site"
FT W09746263-A1.
PN 11-DEC-1997.
PD 04-JUN-1997; U05951.
PF 05-JUN-1996; US-659206.
PR (AURA-) AURAGEN INC.
PA Rakhmillevich AL, Yang N;
DR WPI; 98-041898/04.
DR P-PSDB; W44004-05.
PT Interleukin-12 gene therapy of tumours - comprises delivering
PT construct containing promoter and sequences encoding interleukin-12
PT p35 and p40 subunit(s) to target cells in vivo
PS Claim 5; Page 24-30; 50pp; English.
CC Plasmid pWRG3169 is a tandem plasmid encoding both subunits, i.e.
CC p35 (see W44004) and p40 (see W44005), of murine interleukin-12
CC (IL-12). Each subunit gene was cloned from a mouse spleen cDNA
CC library and has been placed under the transcriptional control of
CC a separate cytomegalovirus (CMV) promoter. An SV40 splicing
CC donor/splicing acceptor is provided between each subunit gene and
CC its CMV promoter. The backbone of the plasmid is from pUC19.
CC pWRG3139 induces at least twice the expression of IL-12 as the
CC bicistronic vector pWRG3196 (see V02043) in vivo and in vitro. A
CC novel method of treating tumours in a mammal involves delivering
CC copies of an expressible foreign genetic construct, especially
CC pWRG3169 or pWRG3196, comprising a promoter operative in the
CC mammalian epidermal cells and DNA sequences encoding p35 and p40
CC subunits of IL-12 to target cells in vivo. Delivery of the
CC construct allows IL-12 expression for treatment of solid,
CC metastatic or disseminated tumours, and regression of established
CC tumours. The treatment is effective even when the genetic
CC construct is delivered to a site distant from the tumour.
CC Sequence 7287 BP; 1798 A; 1877 C; 1846 G; 1766 T;
SQ
```

```
Query Match 11.0%; Score 231; DB 1; Length 7287;
Best Local Similarity 99.6%; Pred. No. 2e-41; Indels 1; Gaps 1;
Matches 242; Conservative 0; Mismatches 0;

QY 1 agtgagcgcaacgcaattatgtagtttagctcactcattaggcaccacaggtttacac 60
|||||
DB 7024 AGTGAGCGCAACGCAATTATGTGAGTTAGCTCACTATTAGGCACCCAGGCTTTACAC 7083
|||||

QY 61 tttatgttcgggtcgtatgttggtaattgtgagcggtatacaatttcacacaga 120
|||||
DB 7084 TTTATGTCTTCGGGCTCGTATGTTGTGGAATTGTGAGCGGATAACAATTTTCACACAGA 7143
|||||

QY 121 aacagctatgaccatgattacgccaagctcgaataatcaacctcactaagggaaacaaaag 180
|||||
DB 7144 AACAGCTATGACCATGATTACGCCAAGCTCGAATTAACCTCACTAAGGGAACAAAAG 7203
|||||

QY 181 ctggagctcca-cgcggtggcgccgtctctagaactagtgatccccgggctgcagaa 239
|||||
DB 7204 CTGGAGCTCCACCGGCTGGCGCGCTCTAGAACTAGTGGATCCCGGGCTGCAGGAA 7263
|||||

QY 240 ttc 242
|||
DB 7264 TTC 7266

RESULT 8
V14340/c
ID V14340 standard; DNA; 3699 BP.
AC V14340;
DE 19-MAY-1998 (first entry)
DT Plasmid pBSGFP expressing green fluorescent protein.
KW Green fluorescent protein; GFP; blue fluorescent protein; BFP; jellyfish;
KW genetic engineering marker; gene therapy; plasmid; ss.
OS Synthetic.
PN W09742320-A1.
PD 13-NOV-1997.
PF 07-MAY-1997; U07625.
PR 08-MAY-1996; US-646538.
PA (USSH ) US SEC DEPT HEALTH.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Gaitanaris GA, Pavlakis GN, Stauber RH, Vournakis JN;
DR WPI; 97-558982/51.
PT New nucleic acid encoding proteins of Aequorea victoria with
PT increased fluorescence - useful as markers for detecting cellular
PT transformation, subcellular localisation of proteins, for assessing
PT gene therapy, mutagenicity etc.
PS Example -; Page 67-68; 105pp; English.
CC This sequence is a plasmid used to express the mutated green fluorescent
CC protein (GFP) of Aequorea victoria coding sequence of the invention. The
CC DNA of the invention that encodes a protein that is (a) a protein that
CC has Leu at position 65 (and optionally Thr at 168 and optionally further
CC Cys at 66) and has cellular fluorescence at least 5 times that of
CC wild-type GFP or (b) is a blue fluorescent protein (BFP) with His at
CC position 67 and also at least one of Leu at 65 and Ala at 164, and has
CC cellular fluorescence at least 5 times that of BFP (Tyr67 to His). The
CC nucleic acids can be used as markers in genetic engineering and gene
CC therapy. They may also be used to detect and characterise regulatory and
CC coding sequence elements that control subcellular expression and
CC targeting of proteins. Typical applications are monitoring targeting and
CC transport of proteins in cells; assessment of gene therapy procedures; in
CC diagnosis (when expressed under control of a promoter induced by a
CC particular analyte); assessment of mutagenicity of compounds; and for
CC drug screening (where expression is controlled by the promoter of a
CC target gene), particularly for antiviral or antiparasitic agents.
CC Sequence 3699 BP; 942 A; 884 C; 898 G; 975 T;
SQ
```

```
Query Match 11.0%; Score 230.6; DB 1; Length 3699;
Best Local Similarity 98.0%; Pred. No. 2.1e-41;
Matches 244; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattatgtagtttagctcactcattaggcaccacaggtttacac 60
```

```
|||||
Db 1678 AGTGAGCGCAACGCAATTAATGTAGTGTAGCTCATTAGGACCCCGAGGCTTTACAC 1619
QY 61 ttatgcttccggctcgatgtgtgtgaaattgtgagcgataacaatttcacacagga 120
|||||
Db 1618 TTTATGCTTCCGGCTCGTATGTTGTGTGAATTTGTGAGCGGATAACAATTTTCACAGGA 1559
QY 121 aacagctatgaccatgattacgcgaagctcgaataataacccctcactaaagggaacaaaag 180
|||||
Db 1558 AACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCCCTCACTAAAGGGAACAAAG 1499
QY 181 ctggagctcca-cgcggtggcgccgctctagaactagtgatcccccggtgcaggaa 239
|||||
Db 1498 CTGGAGCTCCACCGCGTGGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA 1439
QY 240 ttccgtgtgc 248
|||||
Db 1438 TTCGATCGC 1430
|||||

RESULT 9
VI3845/C
ID VI3845 standard; DNA; 6046 BP.
AC VI3845;
DE Complete DNA sequence of plasmid pLF120.
KW Plasmid pLF120; canine adenovirus; vaccine; ss.
OS Synthetic.
PN WO9800166-Al.
PD 08-JAN-1998.
PF 30-JUN-1997; U11486.
PR 03-JUL-1996; US-675566.
PR 03-JUL-1996; US-675556.
PA (INMR ) RHONE MERIEUX INC.
PI Fischer L;
DR WPI; 98-086736/08.
PT Canine adenovirus synthetically modified to contain exogenous DNA -
PT where non-essential region of virus has been deleted, useful in
PT immunogenic, immunological or vaccine composition(s)
PS Example 15; Fig 35; 226pp; English.
CC The present plasmid relates to an invention where a canine
CC adenovirus (cad) is synthetically modified to contain exogenous DNA,
CC where a non-essential region of the cad has been deleted.
CC An immunogenic, immunological or vaccine composition comprising the
CC the above cad can be used to induce an immunological response in a
CC host vertebrate, preferably a canine or human, to which it is
CC administered, or transfer genetic information to an animal or
CC human. The exogenous DNA preferably encodes an expression product
CC comprising an epitope of interest, biological response modulator,
CC growth factor, recognition sequence, therapeutic gene or fusion
CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour
CC necrosis factor or melanoma associated antigen.
CC Sequence 6046 BP; 1561 A; 1614 C; 1436 G; 1434 T;
SQ

Query Match 10.9%; Score 229.4; DB 1; Length 6046;
Best Local Similarity 99.2%; Pred. No. 4.3e-41;
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaatataatgtgagtgctcactcattagcaccgccggtttacac 60
Db 3338 AGTGAGCGCAACGCAATTAATGTAGTGTAGCTCATTAGGACCCCGAGGCTTTACAC 3279
QY 61 ttatgcttccggctcgatgtgtgtgaaattgtgagcgataacaatttcacacagga 120
Db 3278 TTTATGCTTCCGGCTCGTATGTTGTGTGAATTTGTGAGCGGATAACAATTTTCACAGGA 3219
QY 121 aacagctatgaccatgattacgcgaagctcgaataataacccctcactaaagggaacaaaag 180
Db 3218 AACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCCCTCACTAAAGGGAACAAAG 3159
QY 181 ctggagctcca-cgcggtggcgccgctctagaactagtgatcccccggtgcaggaa 239
|||||
```

```
Db 3158 CTGGAGCTCCACCGCGTGGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA 3099
QY 240 ttc 242
|||||
Db 3098 TTC 3096
|||||

RESULT 10
VI3844/C
ID VI3844 standard; DNA; 6245 BP.
AC VI3844;
DE Complete DNA sequence of plasmid pLF100.
KW Plasmid pLF100; canine adenovirus; vaccine; ss.
OS Synthetic.
PN WO9800166-Al.
PD 08-JAN-1998.
PF 30-JUN-1997; U11486.
PR 03-JUL-1996; US-675566.
PR 03-JUL-1996; US-675556.
PA (INMR ) RHONE MERIEUX INC.
PI Fischer L;
DR WPI; 98-086736/08.
PT Canine adenovirus synthetically modified to contain exogenous DNA -
PT where non-essential region of virus has been deleted, useful in
PT immunogenic, immunological or vaccine composition(s)
PS Example 14; Fig 33; 226pp; English.
CC The present plasmid relates to an invention where a canine
CC adenovirus (cad) is synthetically modified to contain exogenous DNA,
CC where a non-essential region of the cad has been deleted.
CC An immunogenic, immunological or vaccine composition comprising the
CC the above cad can be used to induce an immunological response in a
CC host vertebrate, preferably a canine or human, to which it is
CC administered, or transfer genetic information to an animal or
CC human. The exogenous DNA preferably encodes an expression product
CC comprising an epitope of interest, biological response modulator,
CC growth factor, recognition sequence, therapeutic gene or fusion
CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour
CC necrosis factor or melanoma associated antigen.
CC Sequence 6245 BP; 1639 A; 1671 C; 1451 G; 1483 T;
SQ

Query Match 10.9%; Score 229.4; DB 1; Length 6245;
Best Local Similarity 99.2%; Pred. No. 4.4e-41;
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaatataatgtgagtgctcactcattagcaccgccggtttacac 60
Db 3537 AGTGAGCGCAACGCAATTAATGTAGTGTAGCTCATTAGGACCCCGAGGCTTTACAC 3478
QY 61 ttatgcttccggctcgatgtgtgtgaaattgtgagcgataacaatttcacacagga 120
Db 3477 TTTATGCTTCCGGCTCGTATGTTGTGTGAATTTGTGAGCGGATAACAATTTTCACAGGA 3418
QY 121 aacagctatgaccatgattacgcgaagctcgaataataacccctcactaaagggaacaaaag 180
Db 3417 AACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCCCTCACTAAAGGGAACAAAG 3358
QY 181 ctggagctcca-cgcggtggcgccgctctagaactagtgatcccccggtgcaggaa 239
Db 3357 CTGGAGCTCCACCGCGTGGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA 3298
QY 240 ttc 242
|||||
Db 3297 TTC 3295
|||||

RESULT 11
VI3843/C
ID VI3843 standard; DNA; 6448 BP.
AC VI3843;
DE Complete DNA sequence of plasmid pLF116.
DT 16-JUL-1998 (first entry)
DE Complete DNA sequence of plasmid pLF116.
```

KW Plasmid pLF116; canine adenovirus; vaccine; ss.  
OS Synthetic.  
PN WO9800166-A1.  
PD 08-JAN-1998.  
PF 30-JUN-1997; U11486.  
PR 03-JUL-1996; US-675566.  
PR 03-JUL-1996; US-675556.  
PA (INMR ) RHONE MERIEUX INC.  
PI Fischer L;  
DR WPI; 98-086736/08.

PT Canine adenovirus synthetically modified to contain exogenous DNA -  
PT where non-essential region of virus has been deleted, useful in  
PT immunogenic, immunological or vaccine composition(s)  
PS Example 14; Fig 31; 226pp; English.  
CC The present plasmid relates to an invention where a canine  
CC adenovirus (CAD) is synthetically modified to contain exogenous DNA,  
CC where a non-essential region of the CAD has been deleted.  
CC An immunogenic, immunological or vaccine composition comprising the  
CC the above CAD can be used to induce an immunological response in a  
CC host vertebrate, preferably a canine or human, to which it is  
CC administered, or transfer genetic information to an animal or  
CC human. The exogenous DNA preferably encodes an expression product  
CC comprising an epitope of interest, biological response modulator,  
CC growth factor, recognition sequence, therapeutic gene or fusion  
CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour  
CC necrosis factor or melanoma associated antigen.  
SQ Sequence 6448 BP; 1740 A; 1698 C; 1482 G; 1527 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6448;  
Best Local Similarity 99.2%; Pred. No. 4.4e-41;  
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaataatgtgagtagctcactcattagcaccgccggtttacac 60  
DB 3740 AGTGAGCGCAACGCAATTAATGTGAGTAGCTCCTACTATTAGGCACCCGAGGCTTTACAC 3681  
QY 61 ttatgtctccgctcgatgtgtggaattgtgagcgataacaatttcacacagga 120  
DB 3680 TTTATGCTTCGGGCTCGTATGTTGTGTGGAAATGTGAGCGGATACAAATTTTCACACAGGA 3621

QY 121 aacagctatgaccatgattacgccaaagctcgaaattaacctcactaaagggaacaaaag 180  
DB 3620 AACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCTCTACTAAAGGAACAAAAG 3561  
QY 181 ctggagctcca-cgcggtgcccgcgtctagaactagtgagatccccgggctgcagaa 239  
DB 3560 CTGAGCTCCACCGCGTGGCGCGCTCTAGAACTAGTGGATCCCCCGGCTGCAGGAA 3501

QY 240 ttc 242  
DB 3500 TTC 3498

RESULT 12  
V13167/c  
ID V13167 standard; DNA; 6580 BP.  
AC V13167;  
DT 16-JUL-1998 (first entry)  
DE Complete DNA sequence of plasmid pLF086.  
KW Plasmid pLF086; canine adenovirus; vaccine; ss.  
OS Synthetic.  
PN WO9800166-A1.  
PD 08-JAN-1998.  
PF 30-JUN-1997; U11486.  
PR 03-JUL-1996; US-675566.  
PR 03-JUL-1996; US-675556.  
PA (INMR ) RHONE MERIEUX INC.  
PI Fischer L;  
DR WPI; 98-086736/08.

PT Canine adenovirus synthetically modified to contain exogenous DNA -  
PT where non-essential region of virus has been deleted, useful in  
PT immunogenic, immunological or vaccine composition(s)

PS Example 6; Fig 7; 226pp; English.  
CC The present plasmid relates to an invention where a canine  
CC adenovirus (CAD) is synthetically modified to contain exogenous DNA,  
CC where a non-essential region of the CAD has been deleted.  
CC An immunogenic, immunological or vaccine composition comprising the  
CC the above CAD can be used to induce an immunological response in a  
CC host vertebrate, preferably a canine or human, to which it is  
CC administered, or transfer genetic information to an animal or  
CC human. The exogenous DNA preferably encodes an expression product  
CC comprising an epitope of interest, biological response modulator,  
CC growth factor, recognition sequence, therapeutic gene or fusion  
CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour  
CC necrosis factor or melanoma associated antigen.  
SQ Sequence 6580 BP; 1710 A; 1790 C; 1514 G; 1565 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6580;  
Best Local Similarity 99.2%; Pred. No. 4.4e-41;  
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaataatgtgagtagctcactcattagcaccgccggtttacac 60  
DB 4696 AGTGAGCGCAACGCAATTAATGTGAGTAGCTCCTACTATTAGGCACCCGAGGCTTTACAC 4637  
QY 61 ttatgtctccgctcgatgtgtggaattgtgagcgataacaatttcacacagga 120  
DB 4636 TTTATGCTTCGGGCTCGTATGTTGTGTGGAAATGTGAGCGGATACAAATTTTCACACAGGA 4577

QY 121 aacagctatgaccatgattacgccaaagctcgaaattaacctcactaaagggaacaaaag 180  
DB 4576 AACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCTCTACTAAAGGAACAAAAG 4517  
QY 181 ctggagctcca-cgcggtgcccgcgtctagaactagtgagatccccgggctgcagaa 239  
DB 4516 CTGAGCTCCACCGCGTGGCGCGCTCTAGAACTAGTGGATCCCCCGGCTGCAGGAA 4457  
QY 240 ttc 242  
DB 4456 TTC 4454

RESULT 13  
V13842/c  
ID V13842 standard; DNA; 6613 BP.  
AC V13842; 1998 (first entry)  
DT 16-JUL-1998 (first entry)  
DE Complete DNA sequence of plasmid pLF102.  
KW Plasmid pLF102; canine adenovirus; vaccine; ss.  
OS Synthetic.  
PN WO9800166-A1.  
PD 08-JAN-1998.  
PF 30-JUN-1997; U11486.  
PR 03-JUL-1996; US-675566.  
PR 03-JUL-1996; US-675556.  
PA (INMR ) RHONE MERIEUX INC.  
PI Fischer L;  
DR WPI; 98-086736/08.  
PT Canine adenovirus synthetically modified to contain exogenous DNA -  
PT where non-essential region of virus has been deleted, useful in  
PT immunogenic, immunological or vaccine composition(s)  
PS Example 13; Fig 29; 226pp; English.  
CC The present plasmid relates to an invention where a canine  
CC adenovirus (CAD) is synthetically modified to contain exogenous DNA,  
CC where a non-essential region of the CAD has been deleted.  
CC An immunogenic, immunological or vaccine composition comprising the  
CC the above CAD can be used to induce an immunological response in a  
CC host vertebrate, preferably a canine or human, to which it is  
CC administered, or transfer genetic information to an animal or  
CC human. The exogenous DNA preferably encodes an expression product  
CC comprising an epitope of interest, biological response modulator,  
CC growth factor, recognition sequence, therapeutic gene or fusion  
CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour  
CC necrosis factor or melanoma associated antigen.

SQ Sequence 6613 BP; 1724 A; 1781 C; 1531 G; 1576 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6613;  
Best Local Similarity 99.2%; Pred. No. 4.4e-41;  
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattatgtgagttagctcactcattagcgaccccgctttacac 60  
|||||  
Db 3905 AGTGAGCGCAACGCAATTATGTGAGTTAGCTCATTAGGACCCCGAGGCTTTACAC 3846  
|||||  
QY 61 ttatgtctccggctcgatgtgtgtggaattgtgagcggaatacaatttcacacagga 120  
|||||  
Db 3845 TTATGCTTCCGGCTCGTATGTTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 3786  
|||||  
QY 121 aacagctatgaccatgattacgccaagctcgaaattaaacctcactaaagggaacaaaag 180  
|||||  
Db 3785 AACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCTCTATAAGGGGAACAAAAG 3726  
|||||  
QY 181 ctggagctcca-cgcggtggtggtggtggaattgtgagcggaatacaatttcacacagga 239  
|||||  
Db 3725 CTGGAGCTCCACCGCGGTGCGGCCCTCTAGAACTAGTGGATCCCGGGCTGCAGGAA 3666  
|||||

QY 240 ttc 242

Db 3665 TTC 3663

#### RESULT 14

Q39050  
ID Q39050 standard; DNA; 6824 BP.  
AC Q39050;  
DE 28-JUL-1993 (first entry)  
DT K.lactis/S. cerevisiae genetic vector.  
KW Genetic; vector; integration; Kluyveromyces lactis; 255 ribosomal DNA;  
KW Saccharomyces cerevisiae; E. coli; domain; yeast; plasmid; promoter;  
KW expression cassette; HIS3; marker; transformant; human; lysozyme; HL2;  
KW GAL7; signal sequence; killer toxin; transcription termination signal;  
KW FLP; 2 micron plasmid; ss.  
OS Synthetic.  
PN EP-537456-A.  
PD 21-APR-1993.  
PF 31-AUG-1992; 114838.  
PR 04-SEP-1991; IT-MI2349.  
PA (ISTS ) SCLAVO SPA.  
PI Galeotti CL, Gallo E, Riccio ML, Rossolini GM, Thaller MC;  
DR WPI; 93-127394/16.  
PT Vector for Kluyveromyces lactis and Saccharomyces cerevisiae -  
PT which allows stable multiple integration of DNA for prodn. of  
PT heterologous proteins  
PS Claim 1; Fig 1; 26pp; English.  
CC This sequence represents a genetic vector which allows the stable  
CC multiple integration of DNA sequences into the genome of Kluyveromyces  
CC lactis and Saccharomyces cerevisiae. This sequence can be used in an  
CC integrating vector which comprises a region necessary for the stable  
CC maintenance of the plasmid in E. coli and a domain which acts as an  
CC integrating unit consisting of two not contiguous sequences of the 255  
CC ribosomal DNA from S. cerevisiae, flanking a genetic marker suitable  
CC for selection of the yeast transformants in which the integration  
CC event has occurred. Other DNA sequences may be introduced into the  
CC integration plasmid, such as expression cassettes. The gene HIS3  
CC from K. lactis and S. cerevisiae is pref. used as a genetic marker  
CC for the selection of transformants and an expression cassette for the  
CC production and secretion into the culture medium of human lysozyme.  
CC This complete transformation vector is 7850 bp long and includes the  
CC integration vector of the invention and an expression cassette  
CC comprising the K. lactis GAL7 promoter, the signal sequence of the K.  
CC lactis killer toxin, the cDNA encoding the ripe form of human lysozyme  
CC (HL2) and the transcription termination signal FLP of the 2 micron  
CC plasmid from S. cerevisiae.  
SQ Sequence 6824 BP; 1815 A; 1521 C; 1726 G; 1762 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6824;  
Best Local Similarity 99.2%; Pred. No. 4.5e-41;  
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattatgtgagttagctcactcattagcgaccccgctttacac 60  
|||||  
Db 6570 AGTGAGCGCAACGCAATTATGTGAGTTAGCTCATTAGGACCCCGAGGCTTTACAC 6629  
|||||  
QY 61 ttatgtctccggctcgatgtgtgtggaattgtgagcggaatacaatttcacacagga 120  
|||||  
Db 6630 TTATGCTTCCGGCTCGTATGTTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 6689  
|||||  
QY 121 aacagctatgaccatgattacgccaagctcgaaattaaacctcactaaagggaacaaaag 180  
|||||  
Db 6690 AACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCTCTATAAGGGGAACAAAAG 6749  
|||||  
QY 181 ctggagctcca-cgcggtggtggtggtggaattgtgagcggaatacaatttcacacagga 239  
|||||  
Db 6750 CTGGAGCTCCACCGCGGTGCGGCCCTCTAGAACTAGTGGATCCCGGGCTGCAGGAA 6809  
|||||  
QY 240 ttc 242

Db 6810 TTC 6812

#### RESULT 15

V13165/c  
ID V13165 standard; DNA; 6959 BP.  
AC V13165;  
DT 16-JUL-1998 (first entry)  
DE Complete DNA sequence of plasmid pLF047A.  
KW Plasmid pLF047A; canine adenovirus; vaccine; ss.  
OS Synthetic.  
PN WQ9800166-A1.  
PD 08-JAN-1998.  
PF 30-JUN-1997; U11486.  
PR 03-JUL-1996; US-675566.  
PR 03-JUL-1996; US-675556.  
PA (INMR ) RHONE MERIEUX INC.  
PI Fischer L;  
DR WPI; 98-086736/08.  
PT Canine adenovirus synthetically modified to contain exogenous DNA -  
PT where non-essential region of virus has been deleted, useful in  
PT immunogenic, immunological or vaccine composition(s)  
PS Example 6; Fig 3; 26pp; English.  
CC The present plasmid relates to an invention where a canine  
CC adenovirus (cad) is synthetically modified to contain exogenous DNA,  
CC where a non-essential region of the cad has been deleted.  
CC An immunogenic, immunological or vaccine composition comprising the  
CC the above cad can be used to induce an immunological response in a  
CC host vertebrate, preferably a canine or human, to which it is  
CC administered, or transfer genetic information to an animal or  
CC human. The exogenous DNA preferably encodes an expression product  
CC comprising an epitope of interest, biological response modulator,  
CC growth factor, recognition sequence, therapeutic gene or fusion  
CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour  
CC necrosis factor or melanoma associated antigen.  
SQ Sequence 6959 BP; 1834 A; 1880 C; 1591 G; 1653 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6959;  
Best Local Similarity 99.2%; Pred. No. 4.5e-41;  
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattatgtgagttagctcactcattagcgaccccgctttacac 60  
|||||  
Db 4264 AGTGAGCGCAACGCAATTATGTGAGTTAGCTCATTAGGACCCCGAGGCTTTACAC 4205  
|||||  
QY 61 ttatgtctccggctcgatgtgtgtggaattgtgagcggaatacaatttcacacagga 120  
|||||  
Db 4204 TTATGCTTCCGGCTCGTATGTTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 4145  
|||||  
QY 121 aacagctatgaccatgattacgccaagctcgaaattaaacctcactaaagggaacaaaag 180

|||||  
Db 414 AACAGCTATGACCATGATTACGCCAAGTCGGAATTAACCCCTCACTAAAGGGAACAAAAG 4085  
Qy 181 ctggagctcca-cgcggtgcccgcctctagaactagtagtggatccccccgggctgcaggaa 239  
|||||  
Db 4084 CTGGAGCTCCACCGCGGTGGCGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAA 4025  
Qy 240 ttc 242  
|||  
Db 4024 TTC 4022

Search completed: August 1, 2000, 18:20:53  
Job time: 20174 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 13:36:13 ; Search time 9137.45 Seconds  
(without alignments)  
411.142 Million cell updates/sec

Title: US-09-147-955-11  
Perfect score: 2105  
Sequence: 1 agtggcgcaacgcaattaa.....acaaaaaaaaaaaaaaaaaaaaa 2105

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 972840 seqs, 892348106 residues  
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0  
Maximum DB seq length: 1000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_om:\*
- 4: gb\_ov:\*
- 5: gb\_pat:\*
- 6: gb\_ph:\*
- 7: gb\_pl1:\*
- 8: gb\_pl2:\*
- 9: gb\_pr1:\*
- 10: gb\_pr2:\*
- 11: gb\_pr3:\*
- 12: gb\_ro:\*
- 13: gb\_sts:\*
- 14: gb\_sy:\*
- 15: gb\_un:\*
- 16: em\_fun:\*
- 17: em\_hum1:\*
- 18: em\_hum2:\*
- 19: em\_in:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_sy:\*
- 29: em\_un:\*
- 30: em\_vi:\*
- 31: gb\_htg1:\*
- 32: gb\_htg2:\*
- 33: gb\_in1:\*
- 34: gb\_in2:\*
- 35: em\_ba1:\*
- 36: em\_ba2:\*
- 37: em\_hum3:\*
- 38: em\_hum4:\*
- 39: gb\_pr4:\*
- 40: gb\_htg3:\*
- 41: gb\_htg4:\*
- 42: gb\_htg5:\*
- 43: gb\_htg6:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1864	88.6	1864	7	AB027455
2	455.4	21.6	1507	7	AB013596
3	455.4	21.6	1594	7	AB000623
4	439.8	20.9	1458	7	AB013597
5	424.2	20.2	2043	7	AB013598
6	303.2	14.4	198944	8	ATCHRIV38
7	303.2	14.4	200576	7	ATFCA0
8	263	12.5	195068	8	ATCHRIV42
9	263	12.5	197419	8	ATCHRIV41
10	263	12.5	205065	7	ATFCA4
11	259.8	12.3	1589	8	ATU81293
12	241.4	11.5	84203	49	AC005106
13	241.4	11.5	103223	8	AC007153
14	232.8	11.1	3481	12	MARIREDM2
15	232	11.0	476	1	CSPAG13
16	231	11.0	2958	14	ARBLSKM
17	231	11.0	2958	14	ARBLSKP
18	231	11.0	3062	14	CVU47947
19	231	11.0	5192	14	PVPJ12581
20	230.4	10.9	1353	5	A67220
21	229.4	10.9	2964	14	SYNBLDKPV
22	229.4	10.9	2964	14	SYNBLSKMV
23	229.4	10.9	6824	5	A25909
24	229.4	10.9	7372	14	SYNPHSCSKV
					AB027455 Petunia x
					AB013596 Perilla f
					AB000623 Nicotiana
					AB013597 Perilla f
					AB013598 Verbena h
					AL161538 Arabidops
					297335 Arabidops
					AL161542 Arabidops
					AL161541 Arabidops
					297339 Arabidops
					U81293 Arabidops
					AC005106 Genomic s
					AC007153 Arabidops
					X68127 M.auratus m
					247172 Calothrix D
					X52324 pBluescript
					X52325 pBluescript
					U47947 Cloning vec
					X98363 Cloning vec
					A67220 Sequence 4
					L08787 Bluescribe
					L08786 Bluescribe
					A25909 Synthetic Y
					L08874 PhageScript

```

c 25 228.2 10.8 4133 14 AF106619
c 26 227.8 10.8 1832 10 D34614 Human TBXAS
c 27 227.8 10.8 2961 14 ARBL2SKM
c 28 227.8 10.8 2961 14 ARBL2SKP
c 29 227.8 10.8 2976 14 CVU61229
c 30 227.8 10.8 3633 14 ASPGREEN2
c 31 227.8 10.8 3886 14 XXU35129
c 32 227.8 10.8 4051 14 XXU35126
c 33 227.8 10.8 4285 14 XXU35137
c 34 227.8 10.8 4480 14 XXU35134
c 35 227.8 10.8 4869 14 XXU35130
c 36 227.8 10.8 4755 14 U93713
c 37 227.8 10.8 4842 14 XXU35125
c 38 227.8 10.8 4893 14 U93714
c 39 227.8 10.8 5035 14 XXU35128
c 40 227.8 10.8 5060 14 XXU35127
c 41 227.8 10.8 5275 14 XXU35135
c 42 227.8 10.8 5407 14 U93715
c 43 227.8 10.8 5434 14 AF016889
c 44 227.8 10.8 5528 14 U93717
c 45 227.8 10.8 5634 14 CVU14120

```

## ALIGNMENTS

```

RESULT 1
LOCUS AB027455 1864 bp mRNA PLN 07-JAN-2000
DEFINITION Petunia x hybrida PH1 mRNA for anthocyanin 5-O-glucosyltransferase,
complete cds.
ACCESSION AB027455
VERSION AB027455.1 GI:6683051
KEYWORDS anthocyanin 5-O-glucosyltransferase.
SOURCE Petunia x hybrida (cultivar:Surfinia) corolla cDNA to mRNA.
ORGANISM Petunia x hybrida
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
REFERENCE 1 (bases 1 to 1864)
AUTHORS Yamazaki, M. and Saito, K.
TITLE Anthocyanin 5-O-glucosyltransferase
JOURNAL Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 1864)
AUTHORS Yamazaki, M. and Saito, K.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) to the DDBJ/EMBL/GenBank databases. Mami
Yamazaki, Chiba University, Faculty of Pharmaceutical Sciences,
Inage-ku Yayoi-cho 1-33, Chiba, Chiba 263-8522, Japan
(E-mail: mamiy@p.chiba-u.ac.jp, Tel:81-43-290-2905,
Fax:81-43-290-2905)

```

## FEATURES

```

source
1..1864
/organism="Petunia x hybrida"
/cultivar="Surfinia"
/db_xref="taxon:4102"
/tissue_type="corolla"
100..1506
/gene="PH1"
100..1506
/gene="PH1"
/codon_start=1
/product="anthocyanin 5-O-glucosyltransferase"
/protein_id="BA89009.1"
/db_xref="GI:6683052"
/translation="MVQPHVLTTPAAGHINPALQFAKNLVKMGIEVTFSTSIYAQS
RNDKSIINAPKLNFPFSDGDFGSHKDPVFMQSKRCSETVKKILITSEN
GQPIICLLYSIFLPAWAEVAREVHIPSAALLWSQAPATILDIYVFNFGHYKAMANSND
PNWSIQLPGLLETRDLPSPFLPYGAKGLRVALPPFKELIDTDAETTPKILVNTF
DELEPEALNAIEGYFYGGPLIPSAFLGNDPLDASFGDLFQNSDMYMWLNKSP
SSVYISGSLNPNISOMEIESKGLIDIGRFLVIVIKENKCKEENKKGCCIELE
KIGKLVPCWSQLELKHPSLGCVFVSHCGWNSALESACCGVPVVAFPQWDMQNAKQV
EDWVKSGRVVRINEDGVVESEIKRIELVMDGKGELRKNNAKKWELAREAVKQEG

```

gene

CDS

```

BASE COUNT 603 a 318 c 369 g 574 t
ORIGIN

```

Query Match 88.6%; Score 1864; DB 7; Length 1864;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1864; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 cegttgctgtgcgacaatttacaaccagaagaataagacatccctttccccccctaaa 301

Db 1 CCgTTGCTGTGCGACAAATTTACAAACCAAGAAATTAAGCATCCCTTTCCGCCCTTAA 60

QY 302 aaacatacaagtttttaatttttttcaactaagcaagaaataatgctgagcctcatgcac 361

Db 61 AAACATACAAAGTTTTTAATTTTTTCACTAAGCAAGAAATATGGTGAGCCCTCATGTCATC 120

QY 362 ttaacaacatttccagacacagaagcccatattatccagcacttccaatttgcagaatctt 421

Db 121 TTAACAACATTTCCAGCACAGAAGCCCATATTAATCCAGCACTTCAATTTGCCAAGATCTT 180

QY 422 gtcaagatgggcatagaagtgcatttttctacaagcattttatgccaaagccgtatggat 481

Db 181 GTCAAGATGGGCATAGAAGTGACATTTTCTACAGCAATTTATGCCCAAGCCGTATGGAT 240

QY 482 gaaaaatccattcttaatgcacaaagagattgaatttcatttcatttcctgagtgcttt 541

Db 241 GAAAAATCCATCTTATGATGCCAAAGGATTGAATTTTCATTTCCATTTCCCGATGGCTTT 300

QY 542 gatgaaggttttgatcattcaaaagaccctgtattttacatgtcacaacttcgcaaatgt 601

Db 301 GATGAAGGTTTTCATCATTTCAAAAGACCCTGTATTTTACATGTGCACAACTTCGTAATGT 360

QY 602 ggaagtgaactgcacaaaaataattctcactgtctctgaaatggagagcctataact 661

Db 361 GGAAGTGAAACTGCAAAAAAATAATTCACCTTGCTCTGANAATGGACAGCCTATAACT 420

QY 662 tgctactttactccatttccctctcttgagcagagagtagcagtgagtgacatc 721

Db 421 TGCTACTTTACTCCATTTTCCCTCTCTGGGAGCAGAGGTAGCAGTGGAAGTTCACATC 480

QY 722 cctctgctctctcttgaggagtcacacagcaacaatatggacacatatattacttcaacttt 781

Db 481 CTTCTGCTCTCTCTTTGGAGTCAACAGCAACAATAATGGACATATATATTACTTCAACTTT 540

QY 782 catggatatgaaaagctatgggttaataatgaatccaatgatcccaatgggtccatcaactt 841

Db 541 CATGGATATGAAAAGCTATGGCTAATGAATCCAATGATCCAATTTGGTCCATTTCAACTT 600

QY 842 cccgggcttcactacttggaactcgagatcttctcttacttcttacttctcttctggtgca 901

Db 601 CCCGGGCTTCCACTACTGGAAACTCGAGATCTTCCCTTCATTTTACTTCTTATGGTGCA 660

QY 902 aaaggagtccttcgagttgcattccaccattcaagaagaattgacacacattagatgct 961

Db 661 AAAGGAGTCTTCGAGTTGCACCTTCACCATTCAAAGATTGATAGACACATTAGATGCT 720

QY 962 gaaacacactcctaagattcttcttgataacatttgatgaattagagcctgagggcactcaat 1021

Db 721 GAAACCACTCCTAAGATTCTTGTAATACATTTTGATGAATTAGAGGCTGAGGCACCTCAAT 780

QY 1022 gcaattgaaggttataagtttttgaattggacgttgacgttctctctctcttcttctgggt 1081

Db 781 GCAATTGAAGTTATAGTTTATGGAATTTGACCGGTGATTCTCTCTGCTTCTTCTGGGT 840

QY 1082 ggaatgaccccttagatgcttctcatttgggtggtgattcttttccaaatccaaatgactat 1141

Db 841 GGAATGACCCCTTTAGATGCTTCTTCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 900

QY 1142 atggaatggttaactcaaacgaagcnaattctcatggtgtgttttatatatcttttggagctta 1201

Db 901 ATGGAATGGTTAAACTCAAAGCCAAATTCATCAGTTGTTTATATATCTTTTGGAGTGCTA 960







[illegible]

Db	1255	GCTATAAGAAGGTATCATTTAAAAGACGAGTGTCAGAAAGTGTATAGATCGTAATG	1314
Qy	1610	gatggagagaagggaattgagaagaatgtctaagaataagaagaattggctt	1669
Db	1315	GGAGATGCTGAGGAGGACAGGAATTGAGAAGAATGCCCAAAATGGAAGATTTGGCT	1374
Qy	1670	agagaagctgtgaagggaagggttgattcttcacacaagaatttaaaggccttttatgatga	1728
Db	1375	AAGGAATCTACGAGGAAAAATAGTTCAATCTAATGTAAATCTTAAGGCCTATGTTAATGA	1433

RESULT 4

LOCUS	ABO13597	1458 bp	mRNA	PLN	20-MAR-1999
DEFINITION	Perilla frutescens PF3R6 mRNA for UDP-glucose:anthocyanin 5-O-glucosyltransferase homologue, complete cds.				
ACCESSION	ABO13597				
VERSION	ABO13597.1 GI:4115560				
KEYWORDS	PF3R6; UDP-glucose:anthocyanin 5-O-glucosyltransferase homologue.				
SOURCE	Perilla frutescens (variety:crispa, cultivar:Shikun) leaf cDNA to mRNA.				

  

ORGANISM	Perilla frutescens				
Eukaryote;	Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; Gentianaceae; Lamiales; Lamiaceae; Perilla.				
REFERENCE	1 (sites)				
AUTHORS	Yamazaki,M., Gong,Z., Fukuchi-Mizutani,M., Fukui,Y., Tanaka,Y., Kusumi,T. and Saito,K. Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin				
JOURNAL	J. Biol. Chem. 274 (11), 7405-7411 (1999)				
MEDLINE	99167509				
REFERENCE	2 (bases 1 to 1458)				
AUTHORS	Yamazaki,M., Saito,K. and Gong,Z.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami Yamazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. & Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan (E-mail:mamiyep.chiba-u.ac.jp, Tel:+81-43-290-2905, Fax:+81-43-230-2905)				

  

FEATURES	Location/Qualifiers
source	1..1458
	/organism="Perilla frutescens"
	/cultivar="Shikun"
	/variety="Crispa"
	/db_xref="taxon:48386"
	/tissue_type="leaf"
gene	29..1360
	/gene="PF3R6"
CDS	29..1360
	/gene="PF3R6"
	/codon_start=1
	/product="UDP-glucose:anthocyanin 5-O-glucosyltransferase homologue"
	/protein_id="BAA36422.1"
	/db_xref="GI:4115561"
	/translation="MVSRVLLATFPAGQHINPALQFAKLLKAGTDVTFTSVYANFRNMNTASAGANGPGLDFVPFSDGYDCLPGKGGRKYMSEMKARGSEALNNLLNDITVTVVYSHLFWAAAEVARISHVPTALLWEPATVICIHYFYNGADEIDAGSNETIQLPRLPQRSLPTFLPATPERFLMMKEKLETLDGEAKVLNVTFDLPEDPALTATIDRYELDIGLPISAFSLDDPSSETSGDGLFEKNVCNEVKSPKSSVYVSFGSVLRFPKAOMEETIGKLACGRPPLAMIREQNKKDDGEEEREELSCIGLKWKMGKGVSWCSQLVELAHLPALGCPFCRHGNNSAVESLCGPVYVPWFQDTTNKLLIEDAGTGVRRVMNEGGSVDCEICERVCMVMDGGDKTLVRNAIKWKTLRAQMG"

  

polya_site	918
BASE COUNT	353 a 348 c 438 g 319 t
ORIGIN	/note="16 a nucleotides"

  

Query Match	20.9%	Score 439.8	DB 7	Length 1458
-------------	-------	-------------	------	-------------

Best Local Similarity	61.0%;	Pred. No. 6.3e-78;	
Matches	855;	Conservative	0;
Mismatches	497;	Indels	57;
Gaps	7;		

337	aaatatggtgagcctcatgtcatcttaatacaacattttccagcaagaagccatattaatcc	396
Qy		
338	aaatatggtgagcctcatgtcatcttaatacaacattttccagcaagaagccatattaatcc	397
Ddb		
25	AAAAATGGTCCGCCGCCGCGTCTGCTAGCAACATTTTTTCGGCGCAAGCGCACATAAATACC	84
Qy		
397	agcacttcaatttgcgaagaactcttgtaagatgggatgaagaagtgcattttctacaag	456
Ddb		
85	CGCCCTCCAAATTCGGCCAGAGAGACTCTTAAAGCGCGCACTGACGTCACTGTTTTCACGAG	144
Qy		
457	caatttatcccaaaacgctatggatgaaaaaatccatttcttaatg-----accaaaaag	510
Ddb		
145	CGTTTATGATGGCGCGCATGGCCAAACACAGCCTCCGCGCTCCGCGMAACCCACCGGG	204
Qy		
511	atgtaatttcattccattttccogatggtgttgatgaaggttttgcattcattcaaaagacc	570
Ddb		
205	CGTCGACTTGTGGCGGTTCTCCGACGGGTACGAGACGGGTGAAGCCCGCGCGCGACGG	264
Qy		
571	tgtattttcacatgtcacaaacttcgtaaatgtggaagtgaaaactgtcaaaaaataattct	630
Ddb		
265	GAAGCGCTACATGTCGGAGATGAAAGCCCGCGCTCCGAGGCCCTTAAAGAACTCTCTCT	324
Qy		
631	cacttgctctgaaaaatgacagcctataacttgcctacttttactccatttttccctcttg	690
Ddb		
325	CAACAACAGCA-----CGTCACCTTTGTCGCTCTACTCCAGCCTCTTTGGCATG	372
Qy		
691	ggcagcagaggtagcaacgtgaagttcacatccctctctgctcttcttttggagtcacacgc	750
Ddb		
373	GGCGCGGAGGTGGCGCGCTTTGCCACGTCCGACGCCGACGCCCTTCTCTGGGTGAGACCGC	432
Qy		
751	asaaatatggacatatattacttcaactttcatggatatgaaaagctatgcttaatga	810
Ddb		
433	CACCGTGTGTCATATACCACCTTACTTCAAGCGGTACGCACACAGAGATCGAGCGCGG	492
Qy		
811	atccaatgatccaaattggtccattcaacttcccggggttccactacttggaaactcagaga	870
Ddb		
493	TTTCCAATGAA-----ATTTCAGCTCCCTCGGCTTCATCCCTGGAGCAGCGAC	540
Qy		
871	tcttctctcatttttaocttcttatggtgcaaaaggaggtcttcgagttgcacttccacc	930
Ddb		
541	TCCTTCGGAGTGTTCGTGTC-----TCGACGCGCGAGAGATTCCGCTTATGA-----	590
Qy		
931	attcaagaatgatagacacattagatgctgaaaccaactcctcaagattcttgtgaatac	990
Ddb		
591	--TGAAGGAGAGCTTGGAACTTTAGACCGGTGAGAGAGCGGAAGATTTGGTGTAAACAC	648
Qy		
991	atttgatgaattagacctgagggcactcaatgcaattgaagggttatagttttatggaat	1050
Ddb		
649	GTTTGATCGTGGAGCCGATGCACCTCAGCGCTATTGATAGGTATGAGTTGATCGGGAT	708
Qy		
1051	tggacctgtgattcctctgtcttcttgggtggaaaaagacctttagatgcttcattgg	1110
Ddb		
709	CGGGCGGTGTATTCCCTCGCGCTCTTGGAGCGGGAAGATCCCTCCGAAAGCTCTTACGG	768
Qy		
1111	tgggtactctttcaaaattcaaatgactata-----tggaaatggttaaaactcaagcc	1164
Ddb		
769	CGGCCATCTTTTCGAAAATTCGGAGGAGAACTACGCTGGAGTGGTTGAACTCGAAGCC	828
Qy		
1165	aaattcatcagttgtttatatctcttttgggagtcataatgaatccatctatttagccaaat	1224
Ddb		
829	GAAATCTTCGGTGGTGTATGTGCTGTTTTGGAGCGCTTTTGAGTTTCCAAAGGCACAAAT	888
Qy		
1225	ggaggagatatcaaaagggttgatagacatbggaaggccgtttttatggtgataaaaga	1284
Ddb		
889	GGAAAGAGATTGGGAAAGGGCTATTAGCTGGCGGAAGGCCCTTTTATGGATGATACGAGA	948
Qy		
1285	aaatgaaaaagc-----aaagaagaagaaataaaaagcttgttgtattgaaga	1335
Ddb		
949	ACAGAGAAATGACGACGGGAAGAGACAGACAGACAGAGATTGAGTTGCATTTGGGGA	1008
Qy		
1336	atttgaaaaaaataggaaaaaatagttccatggtgttccaaacttgaaagtctcaaacatcc	1395
Ddb		

Db	1009	ATTGAAAAAATAGTGTCTGTGTCTGCGAGTTGGGAGGTTCTGGGGCGCACCC	1058
Qy	1396	atctttaggatgttttcttctcatttggaatcagcccttagagagctttagcttg	1455
Db	1069	TGCGTTGGGATGTTTCGTGACGCATTGTGGGTGGAACCTGGCTGTGGAGAGCTTGAGTTG	1128
Qy	1456	tggagtcgcagttgtggcatttctctcaatgagcagatcaaatgcaaatgccaaacaagt	1515
Db	1129	CGGCATTCCGGTGTGGCGGTGCCAGTGCTTTGATCAGACGACGAATGCGAAGCTGAT	1188
Qy	1516	tgaagatgcttggaagagtgagtgaaagtgagagataaaatgaagatgctgttctgaag	1575
Db	1189	TGAGGATCGCTGGGGGACAGGGGTGACAGTCAGAGTGAAGTGAATGAATGAAGGGGTGGGGTTGATGG	1248
Qy	1576	tgaagaaatcaaaagtgattgaattggtaatggatggagagagagaaaggggagaatt	1635
Db	1249	ATGTGAGATGAAGAGGTGTGGAGATGGTGATGGATGGGGGTGCACAGACCAACTAGT	1308
Qy	1636	gagaaagaatgctaagaaatggaagaattggctcagagaaagctgtgaagggaagtggaac	1695
Db	1309	GAGAGAAATGCCATCAATGGAAGACTTTGGCCAGACAAACCCATGGATAGGATGCATC	1368
Qy	1696	ttcacacaagaatttaagggcttttattgatatgtgttgc	1734
Db	1369	TTCACTCAACAATCTCAACGGCTTTCTTCGTCAAGTTGC	1407

## RESULT 5

ABO13598	2043 bp	mRNA	PUN
LOCUS			
DEFINITION	Verbena hybrida HGt8 mRNA for UDP-glucose:anthocyanin 5-O-glucosyltransferase, complete cds.		20-MAR-1999
ACCESSION	ABO13598		
VERSION	ABO13598.1	GI:4115562	
KEYWORDS	HGt8; UDP-glucose:anthocyanin 5-O-glucosyltransferase.		
SOURCE	Verbena hybrida petal cDNA to mRNA.		
ORGANISM	Verbena x hybrida		
	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliopsida; euclcotyledons; asteridae; Gentiananae; Lamiales; Verbenaceae; Verbena.		
REFERENCE	1 (sites)		
AUTHORS	Yamazaki,M., Gong,Z., Fukuchi-Mizutani,M., Fukui,Y., Tanaka,Y., Kusumi,T. and Saito,K.		
TITLE	Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin		
JOURNAL	J. Biol. Chem. 274 (11), 7405-7411 (1999)		

**FEATURES**  
**SOURCE**

```

1. .2043
/organism="Verbena x hybrida"
/db_xref="taxon:76714"
/tissue_type="petal"
26. .1411
/gene="HGT8"
26. .1411
/gene="HGT8"
/codon_start=1
/product="UDP-glucose:anthocyanin 5-O-glucosyltransferase"
/protein_id="BAA36423.1"
/db_xref="GI:4118563"
/translation="MSRAVLVLLATFPQAQHINPALQPAKRLANNADIOVTFETTSVYAHNR
RMSRTAGSGLNINFEVSFDGDDGLQGGDKNYSKMSRGIKALSDTLAANNVQR
KSHITFVTVLSHPAAKAVAREPHLRASALLIRPATVLDTIEFYFNGYSDLEDAQSD
AIIHGGVLVLAORDLSPELLPSTHPRFRSLMKKLETLLEGEEKPKVLVNSFDEALPD

```







# AUTHORS TITLE JOURNAL

EU Arabidopsis sequencing, project.  
Direct Submission  
Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schueller@mips.mpg.de, mayer@mips.biochem.mpg.de  
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@jnc.ac.uk

## COMMENT

On Jun 30, 1999 this sequence version replaced gi:2244747.  
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>  
this fragment has an overlap with AT5GAL at the 3' end.

## FEATURES

Location/Qualifiers  
1..200576  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"  
1..63289  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/clone="BAC TAMU20K8"  
1753..4428  
/gene="dl3000w"  
join(1753..3194,3261..4428)  
/gene="dl3000w"  
/note="similarity to Cf-2.2, Solanum pimpinellifolium"  
/codon\_start=1  
/product="disease resistance Cf-2 like protein"  
/protein\_id="CAB10171.1"  
/db\_xref="GI:2244748"  
/db\_xref="SPTREMBL:O23253"  
/translation="MTITMSLCIFCLNSILVIATKDLCLPQDQDALLEKNEFSIP  
SPDSDMLIQTKARNNTDCCSWGISCDPKTVVVELDGLNSDLNGLRNSSSLF  
RLHQSLDLSYNDUSCLTPDSSGNFYLRVNLGLCNLFGEIPTSLRSLYTLDDL  
SYNDLTDLEIDSMNKLRLVLSITSCFTGKIPSSIGNLTYLTDLSLWYFTGEL  
PDSMNKSLRVNLHRCNFFGKIPTSLGSLNLTDLDSKNEFTSEGPSMSLNL  
TDQLMLNLSLTNVDLSNNOFKAMLPSSMSLSKLEAFDLSGNSFSSTIPSSLEML  
PSLKLDTGTFDSCPLKIGNISSPNLOELYIGENNINGPIPRSLIKVLGSLALS  
FWDTCGYDFSLFQLKSLRSLDLSGINLNTSSSHLPSMMHLILSCNLSQFPKFL  
ENQTSLYLHANSIOGOVPEWLMLPTLSFIASDNKPSGEIPRAVCEIGTLYLSNN  
NFSGIPPCFISINNTLSILHLRNNLSGVIPPEESLHGYLRSLDVGNSLQSQFPLR  
INCYSYQLNVEENRINTFPSLWLSLNLQVLVLRSEFHPGFSPGDSVLTIKGLR  
FDISENRSFGLVPSDFVGVMSVSEFVIDINTPGFTVVDQDSFHKSVLTIKGLN  
MELVSGSEIVKTIIDVSNRLEGDIPESIGILKELIVLNMNNAFTGHPISLSLSLN  
LQSLDLSNRSLGPSIGELGELTFIARNFSYNMLEGPIPGQTQIQSONSSSFAENPG  
LCGAPLQKKCGEEEDKEKKGKGLSWAAAGIYVPLGFLGALGHILITSYKRDFW  
MRIFCSFSLPKK"  
1753..3194  
/gene="dl3000w"  
/number=1  
3195..3260  
/gene="dl3000w"  
/number=1  
3261..4428  
/gene="dl3000w"  
/number=2  
complement(5905..6149)  
/number=1  
5905..7913  
/gene="dl3000w"  
complement(join(5905..6149,6243..6344,6436..6826,  
7236..7913))  
/note="strong similarity to glycine  
hydroxymethyltransferase (EC 2.1.2.1); contains EST  
gb:1565775"  
/codon\_start=1  
/product="hydroxymethyltransferase"  
/protein\_id="CAB10172.1"  
/db\_xref="GI:2244749"  
/db\_xref="SPTREMBL:O23254"  
/translation="MEPVSSWGNSTSLVSDPVEIHDLEKRRQCRGIELIASENFTS

FAVIEALGSALTNYSEGIPOGNYRGNEFIDEIENICRSRALEAFHCDPAAGVNVQ  
PYSGSPANFAAYTALLOPHDRIMGLDLPSSGHLTHGYTSGKKISATSIYFESLPYK  
VSTFGYIDYDKLEKALDFRPKLLICGGSATPRMDWYARFRAIADKVGALLCDDMAH  
ISGLVAAQEAANFEYCDVVTTHKSLRGPAGMIFYKGPKPQKQEGAGVYDFE  
DKINFAPFALQGGPHNQIGALAVALKQANTPGFYAKQVAKANAVAGLVMKSGY  
QIVTNGTENVHLVLDRLPLGTGNKVKLCDLCSITLKNNAVFGDSSALAPGCVRIGA  
PAMTSGRIVEKDFEIOIGFELSRAVTLTDIQTKYGLKLLKDFNKGLVNNKDLQDLKADY  
EKFSASYEMPGFLSEMKYK"  
complement(6150..6242)  
/number=1  
complement(6243..6344)  
/number=2  
complement(6345..6435)  
/number=2  
complement(6436..6826)  
/number=3  
complement(6827..7235)  
/number=3  
complement(7236..7913)  
/number=4  
12823..14568  
/gene="dl3010w"  
join(12823..13533,13822..14568)  
/gene="dl3010w"  
/note="strong similarity to adenosylhomocysteinase (EC  
3.3.1.1)"  
/codon\_start=1  
/product="adenosylhomocysteinase"  
/protein\_id="CAB10173.1"  
/db\_xref="GI:2244750"  
/db\_xref="SWISS-PROT:O23255"  
/translation="NALLVEKTSRGREYKVKDMSQDFGRLELELAEMPGLMACRT  
EFGSQPKGARITGSLHMTQTAVLIETALGAERKSCNIEFTODHAAATARD  
SNAVFAWKEGTELQEWKCTERDALDNGPGGGDLIYDDGDATLLVHEGVKAEIEFT  
GVGPDPTSDNPEFOIVLSIIEKGLQVDPKPKRKRKRLVGVSEETTCGKRLYQMQQ  
NGTLFPAIVNDSYTKSKFNALYGRSLPDGLMRATDMVMTAGVATCGYDYGKG  
CAAAKFTAGARVITEIDPICALQALMEGLQVLTLEDVVSADIVFTTGNKDIIMVD  
HMRKMNACIVCNIGHFONEIDMLGETVPGVKRITIKPQDRWVPFETKAGIIVLAE  
GLMNLGATGCHPFSMCSPTNOVIAOLELWNEKASCKYEKKVYVLPKHLDEKVAL  
HLGKLGARLTKLSKDQSDYVSIPIEGPKPPHYR"  
12823..13533  
/gene="dl3010w"  
/number=1  
13534..13821  
/gene="dl3010w"  
/number=1  
13822..14568  
/gene="dl3010w"  
/number=2  
complement(16093..16270)  
/number=1  
16093..16501  
/gene="dl3015c"  
complement(join(16093..16270,16317..16501))  
/note="weak similarity to Walleye dermal sarcoma virus gag  
polyprotein"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB10174.1"  
/db\_xref="GI:2244751"  
/db\_xref="SPTREMBL:O23256"  
/translation="MFNSTALVIFAILFILLISADAFPIPSNGEIDAMLIRNSIIGED  
EDLMTEISRRVLMQKRNQELLTIVLALDKLLTAEAVTLPDALETQTTTTHGKI  
QBIKIGSFIAESKLYD"  
complement(16271..16316)  
/number=1  
complement(16317..16501)  
/number=2  
complement(join(18573..18914,18997..19137,19212..19554,  
19603..19719,19782..20123))  
/note="weak similarity to fusion protein la/lb - sugar  
beet"  
/codon\_start=1





Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
1 (bases 1 to 195068)  
EU Arabidopsis sequencing, project.  
AUTHORS  
TITLE  
JOURNAL

Direct Submission  
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: leonke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de  
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,  
E-mail: michael.bevan@bbsrc.ac.uk  
COMMENT  
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/chal/>  
this fragment has an overlap with ATCHRIV41 at the 5' end and an overlap with ATCHRIV43 at the 3' end.

FEATURES  
Location/Qualifiers  
1..195068  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"  
/number="4"  
9917..10066  
/gene="AT4g15560"  
/number=1  
join(9917..10066,10350..10475,10559..10874,10981..11252,11340..11613,11691..11998,12081..12362,12441..12725,12813..12953)  
/gene="AT4g15560"  
/note="GT/AC sites ignored to preserve continuity; light inducible gene required for chloroplast development; Contains Prokaryotic membrane lipoprotein lipid attachment site AA452-462; Prokaryotic membrane lipoprotein lipid attachment site AA591-601; Transketolase signatures AA102-121; Transketolase signatures AA505-521; contains EST gb:AW004332.1, AI997953.1, W43562, N96015, W43388"

exon

CDS

/codon\_start=1  
/product="DEF (CLAL) protein"  
/protein\_id="CAB78598.1"  
/db\_xref="GI:7268304"  
/translation="MASSAFAPSYIITKGLSTDSCKSTLSRSSRLVTDLPSPCLK PNNSHNRRAKVCASLAKEGYYSNRPTPLDITNYP IHKNLSVKELQSLDEL R SDVITFNATGSLGSSGVGLVTHLYIFNTPODKILWDVGHQSYPHKILTGRGK MPTMQTNGLSGFTKRGSEHPCFCTGSHSTISAGLGMVAGRDILGKNNVAVIGD GAWTAQAYEANNAGYLDSMDVILNDKQVSLPTATLDGSPVPVGLSSALSRLQS NPALRELYAKGWTQIGGPMHQLAAKVDYARGMISGTSLSLFEELGLYIIGPVDG HIDLVAALKVKSTRITGTVPLIHVVTKEGRGYPYEARADDDYGVVVFDPATGRF KTNKTQSYTTYFABALVAEVDKDVVAIHAAMGGGTGLNLFQRRFPTRCDFVIAE QHAVTFAAGLAGEGLKPCAITYSSFMQRAYDQVVDVLDQLPVRFBMDRAGLVGADG PTHCGAFDVTMACLPNMIVMAPSDEADLFNNVATAVAIDDRPSCFRYPGNGIGVAL PPGNGVPTIEIGKGRILKEGVALVIGYSAVQSCIGRAVMLEERGLNVTVDARECK PLDRALISLAKSHEVLITVEGSGTGGRGSHVYQFLADGLLDGLKLRWVPLPORYI DHGAPADQLAEAGLMPHSHTAATNALNLIGAPREAL"

gene

intron

exon

intron

exon

intron

exon

/number=4  
11253..11339  
/gene="AT4g15560"  
/number=4  
11340..11613  
/gene="AT4g15560"  
/number=5  
11614..11690  
/gene="AT4g15560"  
/number=5  
11691..11998  
/gene="AT4g15560"  
/number=6  
11999..12080  
/gene="AT4g15560"  
/number=6  
12081..12362  
/gene="AT4g15560"  
/number=7  
12363..12440  
/gene="AT4g15560"  
/number=7  
12441..12725  
/gene="AT4g15560"  
/number=8  
12726..12812  
/gene="AT4g15560"  
/number=8  
12813..12953  
/gene="AT4g15560"  
/number=9  
14413..14436  
20338..24557  
/gene="AT4g15570"  
20338..20373  
/gene="AT4g15570"  
/number=1  
join(<20338..20373,20767..20852,20962..21012,21256..21367,21467..21572,21653..21750,21989..22117,22182..22274,22625..22678,22957..23015,23122..23206,23286..23765,24183..24289,24366..24557)  
/gene="AT4g15570"  
/note="similarity to SEN1 protein, yeast; Contains ATP/GTP-binding site motif A (P-loop) AA13-20; contains EST gb:N65040"  
/codon\_start=1  
/product="SEN1 like protein"  
/protein\_id="CAB78599.1"  
/db\_xref="GI:7268305"  
/translation="VGLSRKSFVLIQGPPGTGTOTILSILGAIIMHATPARVQSKGTD HEVKKGIQMTIOEKYNHMGRASPWILGVNPDRAIMPEDGDDGDFPTSGNELKPEVYNA SRKYRLVLVCAPSNSALDEIVLRLSSCLRDENAOQVTPKIVRIGLKAHHSVASVL DHLVAKRGSAIDKPKQGTGTGDDISIRTAIRLEEARIVMGFNRINVFATLSFGSGL LAKSRGDDVVIIDEAAQAVEPATLIPTATRCQKQVFLVGDPKQLPATVISTVAQDSGY GTSMFERLQKAGYPVKMLKTOYRMHPELIPISQTKPKIVVSLILTPMIASYFTLQIT IF LTENQIRSFPSKQFVEGALDGDSDTEAOTTRDWHKYRCFPGCFDIDHEGKSOHPGA TGSRYNLDEVEFVLLIYHRLVTMYPKLSSOLAIIISPNYQVTKFKDKRMFGTEA EKVDINTYDGFQVVGSAATLAKSDPLKKNLIESAQORNLFKVSKPLNNFSEENLET MKLTDEMEIPDAFLDESLPVPATYGGDDDFGDDQDDQDDVAMAGED"

intron

exon

intron

exon

intron



Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis

1 (bases 1 to 197419)

EU Arabidopsis sequencing, project.

Direct Submission

Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk

Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/> this fragment has an overlap with ATCHRIV40 at the 5' end and an overlap with ATCHRIV42 at the 3' end.

#### FEATURES

##### source

1. .197419

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db\_xref="taxon:3702"

/chromosome="4"

21638..23498

/gene="AT4g15240"

complement(21638..22378)

/number=1

complement(join(21638..22378,22677..23498))

/note="similarity to hypothetical 66.0 kd protein

F47D12.7; contains EST gb:AA712915"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="CAB78566.1"

/db\_xref="GI:7268271"

/translation="MNQTRKRPPSSLRSLDLSPEMLFPKPSLTSRRINFL

LIFCFIIFSYGNFRREQISSIARSVSFSTRHLLFSIAASHDSLRRSSVVR

LYSPSTRAVFLDRGLESLTLPPVIVDSRFPYFPGSLRSARVARVVKET

VDRGKDVWFVGDDTVFFDNLVTVLSKYDHRKVFYVGSSEFYDQNVRSFDM

FGGGFAISAKLAKVLKLDLSCLMRYSHMGSDSRIFSCVAELGVTLTPEGFHID

VGNLFLGLALPLVSLHLDAVDPEFPRNRTESVAHLIGASFDGSLQGSV

CDLSLNTVTVVWGVAVQYEGNKLDPOLLTLQKTFSTWRGSGVGSNMFSEIYP

RDPGRPLVFFLDVSGSGTGSNINVRHGVRABAEVRLERIRVLSPLERNR

QCMPEDLIAAMT"

/complement(22379..22676)

/number=1

complement(22677..23498)

/number=2

28265..29226

/gene="AT4g15250"

complement(join(28265..28651,28738..29226))

/note="similarity to CONSTANS protein, Arabidopsis

thaliana"

/codon\_start=1

/product="CONSTANS like protein"

/protein\_id="CAB78567.1"

/db\_xref="GI:7268272"

/translation="MEARCDFCGTEKALIIYCKSDSAKCLNCNDVNVHSANPLSQRHTR

SLICEKCSQPTAVHCHMENVSLLCCQCMQNTASCTGLGRLKQSLNPDYSCPSDFGK

INWSTLEPVSILVSPFSDTLIQELDDWSSSTVYVQTQNLKQSDYSSFFPMESNLPK

IEECSGLDCEGILNDAPLNFANDIIGSSLDNTKCYEYDSEKFEENNIGLPSL

LLPTLSGNVVPNKSLSMSNLTGESNATDYDQCGISPGFLIGDSPNESNVEVSFNPKLR

DEAKRYKOKSKRM"

complement(28265..28651)

/number=1

complement(28652..28737)

/number=1

complement(28738..29226)

/number=2

31839..31950

/gene="AT4g15260"

/number=1

31839..33264

##### CDS

/gene="AT4g15260"  
join(31839..31950,32018..33264)  
/gene="AT4g15260"  
/note="strong similarity to UTP-glucose  
glucosyltransferase; Contains UDP-glucosyltransferases  
signature AA319-362; Contains EST gb:T20820, T46737,  
A1993247.1, A1994260.1, A1996826.1, A1997955.1, T76043,  
AA720097, A1997288.1, W43072"  
/codon\_start=1  
/product="glucosyltransferase"  
/protein\_id="CAB78568.1"  
/db\_xref="GI:7268273"  
/translation="MKKIELFIPSPGIGHLRPTVKLAKQLIGSEHFHNSQDRLR  
YPSISVEKOPTANLEPSOVIYEKOKPQVRDVSRIINPTKRLAGVVDVDMFCSSMIDI  
ANEFGPCVMIYTSNATLGIITLHVQEWDDKYYDSDLDSENELEFCLRPYPVK  
CLPHILLSKDWLFFFAQGRSRFRKKGILNVYVALEPHALKNFNVDLPQAYVPGPV  
LHLNDGDDDDRLLEVLRLWDLPDPKPSVLFLFCGSGGGFTTEETQVAVALENRSGRF  
LMSLRASPNIMRPGDYKNLEEVLPDGLFRTLDRGKVIQNAIPQAVALEKPAIGGF  
VTHCGNSMLESILWFCVPMVTPLYAEQVNAFEMVEELGLAVEIRKCISSGOLLIGE  
MEIVTAEDIERAIRCVMEDSDSVRSRKEMAEKCHVALMDGGSSKTALQFIQDVLEN  
VA"  
31951..32017  
/gene="AT4g15260"  
/number=1  
32018..33264  
/gene="AT4g15260"  
/number=2  
34599..34620  
complement(34923..35221)  
/number=1  
34923..35916  
/gene="AT4g15270"  
complement(join(34923..35221,35280..35916))  
/note="similarity to UTP-glucose glucosyltransferase;  
contains EST gb:T20820"  
/codon\_start=1  
/product="UTP-glucose glucosyltransferase like protein"  
/protein\_id="CAB78569.1"  
/db\_xref="GI:7268274"  
/translation="MKIQLVFIPSPGIGHLRSTVELAKRLVRSDDRLWITVLIIPPS  
ISDDVPTVITIASLTDSDRLNIEAISVANQPTDYQPTQVIEKQKQKRDVAVARI  
FISTGVDSPRVAGVVDVDMFCSSMIDVNEFVPCVMYVTSNATCIQLIHQRMEDEK  
YVDSLEDSVNELEFPFLTRPYPVKCLPDFTSKDLAFFLAQAARCFRKKMKVVTWPL  
YAEQKISAFAMVEELGVAQIRKFFRGDMLVGMELVTVTDITERAVRCVMEVNDSEVRN  
RVKEMAEKCHVASMDGSSQVALQKFIQDVENVV"  
complement(35222..35279)  
/number=1  
complement(35280..35916)  
/number=2  
37302..38738  
/gene="AT4g15280"  
/number=1  
37302..38738  
/gene="AT4g15280"  
/number=2  
37302..38738  
/gene="AT4g15280"  
/note="strong similarity to UTP-glucose  
glucosyltransferase; Contains UDP-glucosyltransferases  
signature AA346-389; Contains EST gb:T20820, T46737,  
A1993247.1, A1996826.1, T76043, AA720097, W43072"  
/codon\_start=1  
/product="UTP-glucose glucosyltransferase"  
/protein\_id="CAB78570.1"  
/db\_xref="GI:7268275"  
/translation="MKIELFPLPGIGHLRPTVKLAKQLIGSENRLSITIIIPSRF  
DAGDSACIASLTTLSQDDRLHYESISVAKQPTSDPVPQAVYIEKQTKVRDAVA  
ARIVDPTKRLAGVVDVDMFCSSMIDVNEFVPCVMYVTSNATFLGLMLHQVYDQKK  
YDVSELENSVTELEFPLTRPSVKNCLPHILTSKWLPLSLAQARCFRKKMKGLVNTV  
AELEPHALKWENINGDLPQVYPVGLHLENGDDDEKQSETLRGLWDQPSKSVYFL  
CQSGSLGGFTTEQTRETAVALDRSGORFLACLRHASPNIKTDRPDVTLNLEVPPEGL  
ERTLDGRKVIQNAIPQAVALEKPAIGGFVTHCGNSLSLESLWFCVPMVTPLYAEQVNA  
AFEMVEELGLAVEIRKYLKGLDFAGEMETVTAEDIERAIRRVMEQSDSVNRNNKMAE  
KCHVALMDGGSSKAALKERFIQDVLENM"

##### Intron

##### exon

##### repeat\_region

##### exon

##### gene

##### CDS

##### Intron

##### exon

##### exon

##### gene

##### CDS



Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 205065)

Bevan, M., Stickens, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N., Kreis, M., Kavanagh, T., Entlan, K.D., Rieger, M., James, R., Puigdomenech, P., Hatzopoulos, P., Obermaier, B., Duesterhoft, A., Jones, J., Palme, K., Ansong, W., Delsen, M., Bancroft, I., Mewes, H.W., Schueller, C. and Chalwatzis, N.

Unpublished

2 (bases 1 to 205065)

EU Arabidopsis sequencing, project.

Direct Submission

Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG, E-mail: schueller@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk

On Jun 30, 1999 this sequence version replaced gi:2244901.

Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/> this fragment has an overlap with ATPCA3 at the 5' end and an overlap with ATPCA5 at the 3' end.

Location/Qualifiers

1..205065  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"

source

1..14029  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/clone="cosmid YUP5B11.B1G15/YUP5B11.B1G19"  
4257..7108

gene

4257..4662

exon

4257..4662

CDS

Join(<4257..4662,6078..7108)  
/gene="dl3766w"  
/note="strong similarity to fatty acid hydroperoxide lyase  
- Capsicum annum; contains EST gb:949688"

/codon\_start=1

/product="hydroperoxide lyase (HPOX) like protein"

/protein\_id="CAB45989.1"

/db\_xref="GI:5281016"

/translation="NISAATIPPLTASPPNAGIVRLAVGWTIIIGFFRQGPDKFFRT  
RAEKYKSTVFTNIPPTFFGNNVNPVAVLDVKFSFHLFMDLDVDRDLIGDFRP  
SLGFYGGVGVYLDPTTEPKHAKVTVYALLRLRAMELTKRSSKWLQELRSNLIF  
WGTISESKNGAASYIPLQRCIFSLCASLGVADSVSPDIANGKWTINLWALQ  
VPIPTAKGVVPOPLEIILLHTWPPSLIAGNYKKLYNFIDENAGDCIRLQGEFGLT  
RDEAQNLLFVFNAYGGFVFLPSLGRITDNGSLQERITRVRVCGSDNLF  
KTVNEMELVSVVYELAFSPVLPQAFARAKDFQISSHDVFEVKKGELLGCGYPLV  
MRDANVEPEEFKPDYVGTGSELLANLYWSNGPQTPTPSANKQCAARDIVTLTA  
SLIVADLFLRYDTITGDSGSIKAVYKAK"

intron

4663..6077

/gene="dl3766w"

/number=1

exon

6078..7108

/gene="dl3766w"

/number=2

exon

8083..8712

/gene="dl3768w"

/number=1

gene

8083..9807

/gene="dl3768w"

Join(8083..8712,8790..8899,9225..9395,9521..9574,  
9654..9807)

CDS

/gene="dl3768w"

/note="predicted"

/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB45990.1"  
/db\_xref="GI:5281017"  
/translation="MSFFRSKTSKTETQOPSSYGOQDAMYSSTVTSDEVLLQIHGC  
RAHLNGSEAVELAAAGDFELVQDLSNVALAMVVRIGNDLQMPVIKDPVVKLDSRDY  
LFTLVKDGEPISYGITFFPDIDNVFNSTLELDLDFLRNSCFSSPSSSSSSSS  
SSSVNNGIDMKFAPKIEDYNNVAKAIGTGTGHIIRGFMKCSNAYTNQVHKGGEIM  
ITKAEKNGACSSKRATTNKQINKLQVRKLSRATEKLSKTMKLVGVVSGSYMG  
VYKSPGKAFSPWPFVGLASLDALSRETHSFYQSSSTRVSRLESAGEATKDV  
LGTGHAAGTAWNVEINIRKAFSPSSVTSGLKNASRK"  
8713..8789  
/gene="dl3768w"  
/number=1  
8790..8899  
/gene="dl3768w"  
/number=2  
8900..9224  
/gene="dl3768w"  
/number=2  
9225..9395  
/gene="dl3768w"  
/number=3  
9396..9520  
/gene="dl3768w"  
/number=3  
9521..9574  
/gene="dl3768w"  
/number=4  
9575..9653  
/gene="dl3768w"  
/number=4  
9654..9807  
/gene="dl3768w"  
/number=5  
complement(10065..10511)  
/note="similarity to L.esculentum Tfm5 gene"  
/codon\_start=1  
/product="glycine-rich protein like"  
/protein\_id="CAB10324.1"  
/db\_xref="GI:2244903"  
/db\_xref="SPTREMBL:O23398"  
/translation="NKITKRAHFVNTAAKLAVLLIITLLVNCITVYVAKGGGGAH  
GGGVHVSVGGHAHSGVGGHSGGSHASGGGHAAGGGHAGGGHAGGGGGG  
EKGHGIGRGGMVHRPATRNGSGTLTIPFDVGSSTVFMIIFFLN"  
complement(10065..10511)  
/number=1  
10065..10511  
/gene="dl3770c"  
Join(12306..12597,13233..13495,13554..13835,13944..14000,  
15045..15290,15396..15802,16095..16280,16380..16453,  
17015..17346)  
/gene="dl3775w"  
/note="similar to hypothetical protein C02F5.7,  
Caenorhabditis elegans; contains EST  
gb:1217311:937779:315469"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB10325.1"  
/db\_xref="GI:2244904"  
/db\_xref="SPTREMBL:O23399"  
/translation="WDKPYGVASVSMGIDRSAGKDIDLEMGVGEATLYPGLSYGENO  
LRWGFIRVYGLISQALLTLTLLISAVVNLNPNVDLLGSGPGLFLCIVPILLISFE  
FVSTYELDFVLSRSTAIYFMSYLLTRFIARHVDYWFNFVYVWPLHIIHQHPVNL  
LLALFTVLSFTVSGCAMEGCLPLNLSYKQELNPETNSKACKSCVFTNLL  
FFLRVLPGRIVLQALILTLYVGLSTAYTFAAKKDFSLGPIFLTSLILVYTS  
FIQFFPLGPTSVAVYGSFSAKRRKRSSSSSKRKLKLTDTKSCAENVESSLT  
DTGLTALANGPPIENSLIWCNPNVSVGLCSLAQCTSLKSLDQCVGVGGGLAAV  
GKFCQLEELNLRCEGLTDYGVLDLVVGCSSKLSIGVAASAKITDLSLEAVGSHCK  
LLEVLTLDSYTHDKGLTAVAGQCHRLANKLQCVSVTDVAFAAKCELEKTSERLALY  
SFQFTDKMRAIGKSKKLLDLSDCYFVSCGKGLAIAHGCKELEKTSERLALY  
TRGFIAGKSCPRKELALYCORIGNSALAEIGKYLKAGTFDKHFGNIDMPLAEL  
GEGCPMLKDLVLSHCHHTDNLNHLVQKLLKTCVHMVYCPGITVATVVSQCPH













```

/gene="F3F20.2"
/notes="Hypothetical protein"
/codon_start=1
/protein_id="AAD30611.1"
/db_xref="GI:4836309"
/translation="MESNGGGEVRRVNLVFLSRSGHVDHPHLRVHLSRNGVFLR
DYKWLADAGDAMPADFSWCKRRYKNGYVMDLDDLLTPIPDNRYVLKGSILL
SSPKEDYINVEKKAWVTRNGGIDAEEKLQKLKLTSEKIQKESVPCSORSTATTSTV
RESTINEEGFLVKDDPTVSGORDGSTENGNDVESGRPSVSTSSSYIKNKS
SSVASHVLNMLKCGGLDNDVAVLPUNKSNGAFGANEDERIQTHQHNAKNSF
EGAMGIIKMETIEFCKPKVAPSKPMPLCSQCKLFPKPKMHSHMLKRGMRKNSA
NNDLMTSNTVPRQQRNRNIPGNPLGHQRVLTTTLE"
/complement(47118..50318)
/gene="F3F20.3"
/complement(join(47118..48467,48761..49149,49186..49302,
49523..49979,50064..50318))
/notes="F3F20.3"
/notes="similar to Na/H antiporter proteins; N-terminal
half of protein is similar to Na/H antiporter proteins"
/codon_start=1
/protein_id="AAD30610.1"
/db_xref="GI:4836308"
/translation="MSSGAPLNVTPNDTIBESRCKIVCDQSLFFKREOKGWESG
STLASSLFFITQLEFANLSYKVLVLYLPLPPFAQILCGLLFSPVLGNTRPII
AHVFPYRFTVLETFANLALYNIFLLGLMDLRMRVITELKPVIIATFGLVALPVG
AFLYLPGNGHPDKIISGCVFWSVALCTNFPDLARILADLLKLRSDMRGTAAVAV
TDLCTWLLVFGFASFSKGTWNKMPFVITTAIFVLLCIFVIRPLGIWTFAPKTVKA
GGVLCGLITDAGCVHSTIGAPFLGLSPHDIIRNMLEKLHDFLGLAPLFYIIC
GLRDIQPNLOFTDPMVYVHCSSFLKIVTVITVLSLPHIPMDRDAIGAALNTKG
TSLVVLNAGRDKALDSPMYTHMIALVMSLVVEPLAFAYKPKKLHAKHRTVQ
KIKGTELRANLACVRLPNVSGITNLLQVSNATKQSPLSVFAIHVELTGRTHSLLI
MNDECKPFRANFDRVRAESDQIAETFEAMEVNDAMTQVITAVSPYATMHEDICVLA
EDKRVCTILPYVHKLTDPGRMEGNSHAEINONVLSHAPCSVGLVDRGMWYRSE
SFRGSMREKMEVFGGDDPDEALSYARWVGQHVIKLTVVRFVPGREALISCKVA
ABYEREKQVDDCEIYEFNEKTHNDSVYIEKVVNDGQDTATIREMDDNSYDIYV
GRGYNSPVTAGLNDWSSPELGTIGDTLASSNFTMHASVLVIOQYSATRKQAAVTA
AAATVMAAGVAGTGNLLESAGGDAKMRDAHEPPMKMGNVEDDEDEDEHQYIHR"
/51395..53210
/gene="F3F20.4"
/Join(51395..52184,52258..53210)
/notes="F3F20.4"
/notes="Similar to hexosaminidase; Similar to
hexosaminidase, gi|1170249, gi|1076942 and others"
/codon_start=1
/protein_id="AAD30612.1"
/db_xref="GI:4836910"
/translation="MLTSLKPHVILIPILFFITLLSPFLSTALPININPKRFLSWPQ
HKAIALSPNFTILAPERQTLASVTRYHNLINSENYSPLIYPPVKLMKRYTLRNLVVT
VTFDSLPLHGVDESQKLSIPISFSAHLHSAWGMARGLETFSQMIWGTSPDLCLP
VGIIYQDSPFLFGRGVLLDTSRNYGVDDIMRTIKAMSANKLNVPHWHTSQSPPLV
LPSEPLAAKSGSLGPDVYTPEDVSKIVQYGEHGVRYLPEIDTPGHGSMGEAYPEI
VTCANFWHPAGKSWERLASEPCTGQLNPLSPKTYEVVKNVQIDVQNPSPSPHGG
GDEVIPGKWKTDPAINSLSSGGLSLLEKINTSLPIVYSQNTVYVWEDVLGDQ
IKADPSVLPEKHEITLQTNWNGPENTKRIYAAGRVIVSSSEFYLDCCGHGVLNDSDI
YDQKSGSGSWCAKPTQWSIYNDIADGLNEERKLVLGCEVALHSEQADSTVIDS
RLWPRASALAESLWSGNRDERGVKRCGEAVDRNLNRYRMVKGITGASPIQPLWCLKN
PGMCTVHGALQDQ"
/53685..55199
/gene="F3F20.5"
/53685..55199
/gene="F3F20.5"
/notes="Hypothetical protein"
/codon_start=1
/protein_id="AAD30626.1"
/db_xref="GI:4836924"
/translation="MSVVNRPVRLVPLSLLSQLKKQKNPVTKLFEAKERFPSYGH

```

```

Query Match 11.5%; Score 241.4; DB 8; Length 103223;
Best Local Similarity 52.5%; Pred. No. 1.5e-38;
Matches 756; Conservative 0; Mismatches 601; Indels 84; Gaps 7;
QY 336 aaaaatggtcgagctcatgtcatcttaaacacatttcacagcaagccattataatc 395
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 19392 AAAAAATGGCGCAACCGCATTTTCTACTGTTAACGTTTCCGGCGCAAGGTCACGTGAACC 19393
QY 396 cagcacttcaatttgcgaagaatcttctgaagatgggcatagaaagtgaattttctacaa 455
||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 19332 CATCTCTCGTGTTCGTCGCGCTCATCAAAACAACACTGGCGCACGTGTAACCTTTCCGCA 19273
QY 456 gattttatcccaagccgctatggatgaaaaatccattcttaagtcaccaaaaggaattga 515
Db 19272 CGTGTCTCTGTCTATTCACCGCTCTATGATCCCAACCAACACACAGCTCGAANAATCTCT 19213
QY 516 atttcattccatttcctgagtggttgaaggttttgaatcattcaaaagacctgtat 575
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 19212 CTTTCTTACTTTCTCGACGGATTCGACGCGGAGTCACTCCAAACACGACGACGCTGC 19153
QY 576 tttacatgtcacaactctgtaaatgtggaagtgaactgtcaaaaaataattctcaatt 635
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19152 AAACCGGTTGGTACACTTCGAACGTAATGGCGATAAAGCTCTATCGGATTTTCATCGAAG 19093
QY 636 gctctgaaaaatggaca---gcctataacttgctactcttactccattttcccttccttgg 692
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19092 CYAATCAGAATGGTGACTCTCCGGTAAGTTGCTTGATCTACACGATTTCTCCCACTGGG 19033
QY 693 cagcagagtagcaactgaagtccaactccctctctctctctctctctctctctctctctct 752
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19032 TTCCAAAGTGGCGGTAGATTTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 18973
QY 753 caatatggacatatattacttcaacttcaacttcaacttcaacttcaacttcaacttcaact 812
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18972 TCGCTTTCACATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18920
QY 813 ccaatgatccaaatgtgtccattccaacttcccgggggtccacttacttggaaactcgagatc 872
Db 18921 -----AGTCCCGAATCTACTCTCTCTCGAAATCCCGGATC 18886
QY 873 ttccttcattttacttcccttctgttgcaaaaggaggtcttgagcttgcacttccaccat 932
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18885 TGCCTTCTTCTCTCCTCACCCTCCCAACACGAA-----CAAAGCCGCAACAGAGTAT 18835
QY 933 tcaagaattgatagacacattagatgtgaaaccactcctcaagattcttctgaaatcat 992
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18834 ATCAGAAGCTGATGGATTTTCTCAAGAAGAAATCTAACCCGAAATAATTCGTCAACACAT 18775
QY 993 ttgataattgagcctgagcactcaatgcaatggaaggttataagtttttatggaattg 1052
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18774 TCGATTTCGCTGGAGCCAGAGTTCTTAACAGCTATTCCGAATATAGAAATGGTGGCAGTTG 18715
QY 1053 gaccgtgattcctctctgtcttcttcttcttcttcttcttcttcttcttcttcttcttct 1112
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18714 GTCTTTTACTTCTCGCGGAGATTTTTCACGGAAGCGAATCAGGTAAAGATTTTATCAAGAG 18655
QY 1113 gtgattcttttcaaaattcaaatgactatggaatggtttaactcaaacgcaaatctcat 1172
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18654 ATCATCAAGTAGTAGT-----TATACACTTTGGTTAGACTCGAAACACAGAGTCTCT 18604
QY 1173 cagttgtttatatcttcttgggagctcaatgaaatccatctattatcccaaaatggagaga 1232
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18603 CTGTTATTATGTTCTTTTGGAAACAATGGTTCAGTTGTCGAAGAACAACAGATAGAGAAC 18544
QY 1233 tatcaaaagggttgatagacataggaagccgtttttatgggtgatacaaaagaaaaatg --- 1289
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18543 TAGCAGAGACTCATAGAGGGGGAAGACCGTCTTGTGGGTATTAACGTGATAAATCTCA 18484
QY 1290 -----aaaaaggcaagaagaaga-----ataaaagcttggttga 1328
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18483 ACNAGAACGGAATAATAGAGGAGAGACAGACAGATTTGAGAAGATAGCTGGTTTATA 18424
QY 1329 ttgaagaattggaaaaaataagtaaaatagttccatttggtttcacaacttggaagttctaa 1388
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18423 GACACGAGCTTGAAGAGTTGGGATGATTGCTCGGTGTTCCGACATAGAGGTTTGA 18364
QY 1389 aacatccatttttaggatgttttttcttcattgttggaatggaaattccagccttagagatt 1448
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18363 GACACGAGCCATGTTGTTTTTGTGCTTGTGAGTCAATTTGTTGGGTGAGCTCATCTACTGAGAGTT 18304

```

ORIGIN	Query Match	11.1%	Score 232.8	DB 12	Length 3481
	Best Local Similarity	96.9%	Pred. No. 1e-36		
	Matches 248	Conservative 0	Mismatches 7	Indels 1	Gaps 1
QY	1	agtgagcgacgcaataatgtgagttgactcactcattagcaccgccaggtttacac	60		
DB	2389	AGTGAGCGCAACGCAATTAATGTGAGTTAGTCTACTCATTAGGCACCCAGGCTTTACAC	2330		
QY	61	tttatgcttcggcgctcgtatgtgttggaattgtgagcggaataacaatttcacacagga	120		
DB	2329	TTTATGCTTCGGGCTGCTATGTTGTGTGGAATCTGAGCGGATAACAATTTACACAGGA	2270		
QY	121	aacagctatgaccatgattacgccaagctcgaaattaacccctcactaaagggaacaaaag	180		
DB	2269	AACAGCTATGACCATGATTAACGCCAAGCTCGAAATTAACCCCTCACTAAGGGAACAAAAG	2210		
QY	181	ctgagctccacgcggtggcgccgctctagaactagtgaactcccccggcgctcgagaat	240		
DB	2209	CTGAGGCTCCA - GCGGTGGGCGCCGCTAGAACTAGTGGATGCCCGGGCTGAGGAAT	2151		
QY	241	tccgttgctgcgcga 256			
DB	2150	TCCGGAATGGGCAA 2135			
RESULT 15	CSPAG13	476 bp	DNA	BCT	04-JAN-1995
LOCUS	Calothrix D253	genomic DNA	(clone AG13).		
DEFINITION	247172				
ACCESSION	247172.1	GI:619085			
VERSION					
KEYWORDS					
SOURCE	Calothrix sp.				
ORGANISM	Calothrix sp.				
REFERENCE	Bacteria; Cyanobacteria; Nostocales; Rivulariaceae; Calothrix.				
AUTHORS	1 (bases 1 to 476)				
TITLE	Robinson, N.J., Robinson, P.J. and Gupta, A.				
	Genomic DNA sequence obtained in the process of generating a random				
	sequence database greater than 10 kb from single sequence runs from				
	random clones of a Calothrix D253 library				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 476)				
AUTHORS	Robinson, N.J., Robinson, P.J., Gupta, A., Bleasby, A.J., Whitton, B.A.				
	and Morby, A.P.				
TITLE	Singular over-representation of HIPI in DNA from many cyanobacteria				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 476)				
AUTHORS	Robinson, N.J.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-DEC-1994) Nigel J. Robinson, Biochemistry and				
	Genetics, The Medical School, University of Newcastle, Framlington				
	Place, Newcastle upon Tyne, NE2 4HH, UK				
REFERENCE	4 (bases 1 to 476)				
AUTHORS	Robinson, N.J., Robinson, P.J., Gupta, A., Bleasby, A.J., Whitton, B.A.				
	and Morby, A.P.				
TITLE	Singular over-representation of an octameric palindrome, HIPI, in				
JOURNAL	DNA from many cyanobacteria				
MEDLINE	Nucleic Acids Res. 23 (5), 729-735 (1995)				
FEATURES	9523777				
	Location/Qualifiers				
source	1. 476				
	/organism="Calothrix sp."				
	/strain="D253"				
	/db.xref="taxon:1187"				
	/clone="AG13"				
	/clone_lib="LZAPCalothrix-LIB"				
BASE COUNT	94 a 119 c 132 g 131 t				
ORIGIN					

Query Match 11.0%; Score 232; DB 1; Length 476;  
 Best Local Similarity 99.6%; Pred. No. 1.8e-36;  
 Matches 243; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattaatgtgagttagctcactcattagggcaacccagcgtttacac 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 300 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTTAGGCACCCGAGGCTTTACAC 241  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 ttatgcttccggctcgatgtgtgtggaattgtgagcgagataacaatttcacacagga 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 240 TTTATGCTTCCGGCTCGTATGTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGA 181  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 121 aacagctatgaccatgattacgccaaagctcgaaattaaacctcactaaagggaacaaaaag 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 180 AACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCTCCTAAAGGGAACAAAAAG 121  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 181 ctggagctcca-cgcggtggcgcgcgctctagaactagtggtatcccccggtgcagggaa 239  
 ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||  
 Db 120 CTGGAGCTCCACCGCGGTGGCGCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAA 61  
 ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||

QY 240 ttcc 243  
 |||||  
 Db 60 TTCC 57

Search completed: August 1, 2000, 13:51:48  
 Job time: 58570 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:19:33 ; Search time 152.01 Seconds  
(without alignments)  
1904.842 Million cell updates/sec

Title: US-09-147-955-11  
Perfect score: 2105  
Sequence: 1 agtgagcgcaacgcaattaa.....acaaaaaaaaaaaaaaa 2105

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	231	11.0	7287	4	US-08-659-206A-1
C 2	230.6	11.0	3699	5	US-08-646-538-6
C 3	227.8	10.8	3792	4	US-08-992-334-1
C 4	227.8	10.8	3792	5	US-08-302-752-1
C 5	227.8	10.8	5234	4	US-08-992-334-2
C 6	227.8	10.8	5234	5	US-08-302-752-2
C 7	227.8	10.8	6722	4	US-08-992-334-3
C 8	227.8	10.8	6722	5	US-08-302-752-3
9	224	10.6	2885	2	US-08-471-496-1
10	224	10.6	2885	3	US-08-894-840-1
C 11	211.6	10.1	10281	4	US-08-816-155B-1
12	193.8	9.2	2415	1	US-07-885-970A-26
13	193.8	9.2	2415	1	US-08-298-687A-26
14	193.8	9.2	2415	1	US-08-298-829-26
C 15	192	9.1	14311	6	PCT-US96-06053-1
C 16	192	9.1	14311	6	PCT-US96-06053-7
17	188	8.9	4435	4	US-08-792-824-1
18	187.2	8.9	801	4	US-08-316-50
19	187.2	8.9	949	3	US-08-713-000-4
20	187.2	8.9	949	4	US-08-975-316-4
21	184.8	8.8	5534	2	US-08-452-267-3
C 22	184	8.7	5707	4	US-08-472-809B-8
C 23	184	8.7	6345	4	US-08-472-809B-7
C 24	183	8.7	4539	1	US-08-119-512-1
C 25	183	8.7	4539	2	US-08-488-015B-1
26	182.4	8.7	4525	3	US-08-613-861-2

27	182.2	8.7	7228	4	US-08-850-049-128	Sequence 128, Appl
28	182.2	8.7	7228	4	US-08-850-049-129	Sequence 129, Appl
29	182.2	8.7	7228	4	US-08-050-478-128	Sequence 128, Appl
30	182.2	8.7	7228	4	US-08-050-478-129	Sequence 129, Appl
C 31	181.8	8.6	4164	1	US-08-204-675-1	Sequence 1, Appli
C 32	181.8	8.6	4164	3	US-08-660-754-1	Sequence 1, Appli
C 33	181.8	8.6	4164	6	US-08-796-364-1	Sequence 1, Appli
C 34	181.8	8.6	4164	6	PCT-US95-02520-1	Sequence 1, Appli
C 35	181.8	8.6	4933	1	US-08-204-675-2	Sequence 2, Appli
C 36	181.8	8.6	4933	3	US-08-660-754-2	Sequence 2, Appli
C 37	181.8	8.6	4933	4	US-08-796-364-2	Sequence 2, Appli
C 38	181.8	8.6	4933	6	PCT-US95-02520-2	Sequence 2, Appli
C 39	181.2	8.6	535	2	US-08-463-115-57	Sequence 57, Appl
C 40	181.2	8.6	535	2	US-08-465-388-57	Sequence 57, Appl
C 41	180.8	8.6	5178	3	US-08-474-169-2	Sequence 2, Appli
C 42	180.8	8.6	6206	3	US-08-474-169-3	Sequence 3, Appli
C 43	179.2	8.5	3681	1	US-07-924-028A-6	Sequence 6, Appli
C 44	179.2	8.5	4792	3	US-08-596-300A-2	Sequence 2, Appli
C 45	179.2	8.5	5042	3	US-08-596-300A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-659-206A-1  
; Sequence 1, Application US/08659206A  
; Patent No. 5922685  
; GENERAL INFORMATION:  
; APPLICANT: Rakhmilevich, Alexander  
; TITLE OF INVENTION: IL-12 Gene Therapy of Tumors  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/659,206A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 110229.91144  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7287 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "Plasmid DNA"  
; IMMEDIATE SOURCE:  
; CLONE: pWRG3169  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: 1..628  
; FEATURE:  
; NAME/KEY: iDNA  
; LOCATION: 629..810  
; FEATURE:  
; NAME/KEY: CDS





;;  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,752  
; FILING DATE: 24-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR93/00248  
; FILING DATE: 12-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 932/03034  
; FILING DATE: 13-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Prout, D. Bruce  
; REGISTRATION NUMBER: 20958  
; REFERENCE/DOCKET NUMBER: C93:31779  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (626) 795-9900  
; TELEFAX: (626) 577-8800  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3792 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: YES  
; IMMEDIATE SOURCE:  
; CLONE: pg+host4  
; US-08-992-334-1

Query Match 10.8%; Score 227.8; DB 4; Length 3792;  
Best Local Similarity 98.8%; Pred. No. 1.1e-42;  
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 agtgagcgcaacgaattaatgtagctcactcattagggcaccgccggtttacac 60  
Db 3758 AGTGAGCGCAACGCAATTAATGTAGTTAGCTCATTAGGCAACCCAGGCTTTACAC 3699  
Qy 61 ttatgtctccggctcgatgtgtggaattgtgagcggaataacaatttcacagga 120  
Db 3698 TTTATGCTTCGGCTCGTATGTGTGGAAATGTGAGCGGATAACAATTTACACAGGA 3639  
Qy 121 aacagctatgaccatgattacgcaagctcgaaataaccctcactaaagggaaacaaag 180  
Db 3638 AACAGCTATGACCATGATTAGCCCAAGCGCCCAATTAACCTCTACTAANGGGAACAAAG 3579  
Qy 181 ctgagctccca-cgcggtggcgccgctctagaactagtgatcccccggtgcagga 239  
Db 3578 CTGGAGCTCCACCGCGTGGCGCGCTCTAGAACTAGTGGATCCCCGGGCTGCAGGAA 3519  
Qy 240 ttc 242  
Db 3518 TTC 3516

RESULT 4  
US-08-302-752-1/c  
; Sequence 1, Application US/08302752  
; Patent No. 6025190  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID  
; NUMBER OF SEQUENCES: 3  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/302,752  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:

;;  
; APPLICATION NUMBER: FR 9203034  
; FILING DATE: 13-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO FR/93/00248  
; FILING DATE: 12-MAR-1993  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3792 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-302-752-1

Query Match 10.8%; Score 227.8; DB 5; Length 3792;  
Best Local Similarity 98.8%; Pred. No. 1.1e-42;  
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 agtgagcgcaacgaattaatgtagctcactcattagggcaccgccggtttacac 60  
Db 3758 AGTGAGCGCAACGCAATTAATGTAGTTAGCTCATTAGGCAACCCAGGCTTTACAC 3699  
Qy 61 ttatgtctccggctcgatgtgtggaattgtgagcggaataacaatttcacagga 120  
Db 3698 TTTATGCTTCGGCTCGTATGTGTGGAAATGTGAGCGGATAACAATTTACACAGGA 3639  
Qy 121 aacagctatgaccatgattacgcaagctcgaaataaccctcactaaagggaaacaaag 180  
Db 3638 AACAGCTATGACCATGATTAGCCCAAGCGCCCAATTAACCTCTACTAANGGGAACAAAG 3579  
Qy 181 ctgagctccca-cgcggtggcgccgctctagaactagtgatcccccggtgcagga 239  
Db 3578 CTGGAGCTCCACCGCGTGGCGCGCTCTAGAACTAGTGGATCCCCGGGCTGCAGGAA 3519  
Qy 240 ttc 242  
Db 3518 TTC 3516

RESULT 5  
US-08-992-334-2/c  
; Sequence 2, Application US/08992334  
; Patent No. 5919678  
; GENERAL INFORMATION:  
; APPLICANT: Gruss, Alexandra  
; APPLICANT: Maguin, Emmanuelle  
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE  
; TITLE OF INVENTION: PLASMID  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christie Parker & Hale, LLP  
; STREET: 350 West Colorado Boulevard, Suite 500  
; CITY: Pasadena  
; STATE: California  
; COUNTRY: United States  
; ZIP: 91105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/992,334  
; FILING DATE: 17-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,752  
; FILING DATE: 24-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR93/00248  
; FILING DATE: 12-MAR-1993  
; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: FR FR92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
;
US-08-992-334-2

Query Match 10.8%; Score 227.8; DB 4; Length 5234;
Best Local Similarity 98.8%; Pred. No. 1.2e-42;
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattaatgtgagtgagtcacactcattagcgacccaggtttacac 60
Db 5200 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCATTAGGCAACCCAGGCTTTACAC 5141

QY 61 ttatgcttcggcgtcgatgtgtggaattgtgagcgagatacaaatctcacagga 120
Db 5140 TTTATGCTTCCGGCTCGTATGTTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 5081

QY 121 aacagctatgaccatgattacgcgaagctcgaaattaaacctcaataagggaacaaaag 180
Db 5080 AACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCTCTACTAAAGGGAACAAAAG 5021

QY 181 ctgagagctcca-cgcggtgagcgccgctctagaactagtgatcccccggtgcaggaa 239
Db 5020 CTGGAGCTCACCAGCGGTGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA 4961

QY 240 ttc 242
Db 4960 TTC 4958

RESULT 6
US-08-302-752-2/c
; Sequence 2, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/302,752
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic)
;
US-08-302-752-2

Query Match 10.8%; Score 227.8; DB 5; Length 5234;
Best Local Similarity 98.8%; Pred. No. 1.2e-42;
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattaatgtgagtgagtcacactcattagcgacccaggtttacac 60
Db 5200 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCATTAGGCAACCCAGGCTTTACAC 5141

QY 61 ttatgcttcggcgtcgatgtgtggaattgtgagcgagatacaaatctcacagga 120
Db 5140 TTTATGCTTCCGGCTCGTATGTTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 5081

QY 121 aacagctatgaccatgattacgcgaagctcgaaattaaacctcaataagggaacaaaag 180
Db 5080 AACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCTCTACTAAAGGGAACAAAAG 5021

QY 181 ctgagagctcca-cgcggtgagcgccgctctagaactagtgatcccccggtgcaggaa 239
Db 5020 CTGGAGCTCACCAGCGGTGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA 4961

QY 240 ttc 242
Db 4960 TTC 4958

RESULT 7
US-08-992-334-3/c
; Sequence 3, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; APPLICANT: Maguin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/992,334
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR FR92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

```

;  
; LENGTH: 6722 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
US-08-992-334-3

Query Match 10.8%; Score 227.8; DB 4; Length 6722;  
Best Local Similarity 98.8%; Pred. No. 1.3e-42;  
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgaattaatgtgagtagctcactcattagggcaccgccaggtttacac 60  
DB 6688 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACAC 6629  
QY 61 ttatgttcggctcgatgttgtggaattgtgagcgagataacaatttcacacagga 120  
DB 6628 TTTATGCTTCGGGCTCGTATGTTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 6569  
QY 121 aacagctatgacctgattacgccaagctcgaaattaaacctcactaaagggaacaaaag 180  
DB 6568 AACAGCTATGACCATGATTAGCCCAAGCGCGCAATTAACCTCACTAAAGGGAACAAAAG 6509  
QY 181 ctgagagtcaca-cgcggtggcgccgctctagaactagtggatccccgggctgcaggaa 239  
DB 6508 CTGGAGCTCCACCGGCTGGCGCGCTCTAGAACTAGTGGATCCCCGGGCTGCAGGAA 6449  
QY 240 ttc 242  
DB 6448 TTC 6446

## RESULT 8

US-08-302-752-3/c  
; Sequence 3, Application US/08302752  
; Patent No. 6025190

; GENERAL INFORMATION:

; APPLICANT: THERMOSENSIBLE PLASMID  
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID

; NUMBER OF SEQUENCES: 3

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/302,752

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 9203034

; FILING DATE: 13-MAR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO FR/93/00248

; FILING DATE: 12-MAR-1993

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 6722 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-302-752-3

Query Match 10.8%; Score 227.8; DB 5; Length 6722;  
Best Local Similarity 98.8%; Pred. No. 1.3e-42;  
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgaattaatgtgagtagctcactcattagggcaccgccaggtttacac 60  
DB 6688 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACAC 6629

QY 61 ttatgttcggctcgatgttgtggaattgtgagcgagataacaatttcacacagga 120  
DB 6628 TTTATGCTTCGGGCTCGTATGTTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 6569  
QY 121 aacagctatgacctgattacgccaagctcgaaattaaacctcactaaagggaacaaaag 180  
DB 6568 AACAGCTATGACCATGATTAGCCCAAGCGCGCAATTAACCTCACTAAAGGGAACAAAAG 6509  
QY 181 ctgagagtcaca-cgcggtggcgccgctctagaactagtggatccccgggctgcaggaa 239  
DB 6508 CTGGAGCTCCACCGGCTGGCGCGCTCTAGAACTAGTGGATCCCCGGGCTGCAGGAA 6449  
QY 240 ttc 242  
DB 6448 TTC 6446

## RESULT 9

US-08-471-496-1

; Sequence 1, Application US/08471496

; Patent No. 5798223

; GENERAL INFORMATION:

; APPLICANT: LI, YI

; APPLICANT: CAO, LIANG

; APPLICANT: ROSEN, CRAIG

; TITLE OF INVENTION: HUMAN AMINE TRANSPORTER

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

; STREET: 1100 NEW YORK AVENUE, SUITE 600

; CITY: WASHINGTON

; STATE: DC

; COUNTRY: USA

; ZIP: 20003-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/471,496

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/US95/02645

; FILING DATE: 01-MAR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: STEFFE, ERIC K

; REGISTRATION NUMBER: 36,688

; REFERENCE/DOCKET NUMBER: 1488.0830001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2885 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 719..2128

US-08-471-496-1

Query Match 10.6%; Score 224; DB 2; Length 2885;  
Best Local Similarity 98.7%; Pred. No. 7.2e-42;  
Matches 235; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgaattaatgtgagtagctcactcattagggcaccgccaggtttacac 60  
DB 6688 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACAC 6629

```
Db 207 AGTGAGCGCAACGCAATTAATGTGAGTTAGTCTCATTAGGACACCCAGGCTTTACAC 266
Qy 61 ttatgtctccgctcgtatgtgtggaattgtgagcggaataacaatttcacagga 120
Db 267 TTTATGCTTCGGCTCGTATGTTGTGGGAATTGTGAGCGGATAACAATTTTCACAGGA 326
Qy 121 aacagctatgaccatgattacgcgaagctcgagaattaaacctcactaaagggaacaaaag 180
Db 327 AACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCCTCATTAAAGGGAACAAAAG 386
Qy 181 ctggagctcca-cgcggtggcgccgctctagaaactagtagtgcaccccggtgcagg 237
Db 387 CTGAGCTCCACCGCGGTGGCGNCCGCTCTAGAACTAGTGGATCCCGCGNCTGCAGG 444

RESULT 10
US-08-894-840-1
; Sequence 1, Application US/08894840
; Patent No. 5859200
; GENERAL INFORMATION:
; APPLICANT: LI, YI
; APPLICANT: CAO, LIANG
; APPLICANT: ROSEN, CRAIG A.
; TITLE OF INVENTION: HUMAN AMINE TRANSPORTER
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,840
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0830000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 719..2128
US-08-894-840-1

Query Match 10.6%; Score 224; DB 3; Length 2885;
Best Local Similarity 98.7%; Pred. No. 7.2e-42;
Matches 235; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 agtgagcgcaacgcaatgaatgtgagttagctcactcattaggcaccgccaggtttacac 60
Db 207 AGTGAGCGCAACGCAATTAATGTGAGTTAGTCTCATTAGGACACCCAGGCTTTACAC 266
Qy 61 ttatgtctccgctcgtatgtgtggaattgtgagcggaataacaatttcacagga 120
Db 267 TTTATGCTTCGGCTCGTATGTTGTGGGAATTGTGAGCGGATAACAATTTTCACAGGA 326
Qy 121 aacagctatgaccatgattacgcgaagctcgagaattaaacctcactaaagggaacaaaag 180
Db 327 AACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCCTCATTAAAGGGAACAAAAG 386
Qy 181 ctggagctcca-cgcggtggcgccgctctagaaactagtagtgcaccccggtgcagg 237
Db 387 CTGAGCTCCACCGCGGTGGCGNCCGCTCTAGAACTAGTGGATCCCGCGNCTGCAGG 444
```

```
Db 327 AACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCCTCATTAAAGGGAACAAAAG 386
Qy 181 ctggagctcca-cgcggtggcgccgctctagaaactagtagtgcaccccggtgcagg 237
Db 387 CTGAGCTCCACCGCGGTGGCGNCCGCTCTAGAACTAGTGGATCCCGCGNCTGCAGG 444

RESULT 11
US-08-816-155B-1/c
; Sequence 1, Application US/08816155B
; Patent No. 5990091
; GENERAL INFORMATION:
; APPLICANT: TARTAGLIA, JAMES
; APPLICANT: COX, WILLIAM I.
; APPLICANT: GETTIG, RUSSELL R.
; APPLICANT: MARTINEZ, HECTOR
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: PINCUS, STEVEN E.
; TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
; TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,155B
; FILING DATE: 12-MAR-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: KOWALSKI, THOMAS J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454310-2990
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-588-0800
; TELEFAX: 212-588-0500
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10281 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-816-155B-1

Query Match 10.1%; Score 211.6; DB 4; Length 10281;
Best Local Similarity 97.8%; Pred. No. 6.3e-39;
Matches 225; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1 agtgagcgcaacgcaatgaatgtgagttagctcactcattaggcaccgccaggtttacac 60
Db 4083 AGTGAGCGCAACGCAATTAATGTGAGTTAGTCTCATTAGGACACCCAGGCTTTACAC 4024
Qy 61 ttatgtctccgctcgtatgtgtggaattgtgagcggaataacaatttcacacagga 120
Db 4023 TTTATGCTTCGGCTCGTATGTTGTGGGAATTGTGAGCGGATAACAATTTTCACACAGGA 3964
Qy 121 aacagctatgaccatgattacgcgaagctcgagaattaaacctcactaaagggaacaaaag 180
Db 3963 AACAGCTATGACCATGATTAGCGCAAGCTCGGAATTAACCCCTCATTAAAGGGAACAAAAG 3904
Qy 181 ctggagctcca-cgcggtggcgccgctctagaaactagtagtgcaccccggtgcagg 229
```



Db 361 AACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCCTCACTAAAGGGAACAAAAG 420  
QY 181 ctggagctccacgcggt 197  
Db 421 CTGGAGCTCCACATGCT 437  
RESULT 14  
US-08-298-829-26  
; Sequence 26, Application US/08298829  
; Patent No. 5620882  
; GENERAL INFORMATION:  
; APPLICANT: John, Maliyakal E.  
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON  
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady  
; STREET: P.O. Box 2113, First Wisconsin Plaza  
; CITY: Madison  
; STATE: Wisconsin  
; COUNTRY: USA  
; ZIP: 53701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/298,829  
; FILING DATE: 19-OCT-1994  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/885,970  
; FILING DATE: 18-MAY-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/617,239  
; FILING DATE: 21-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/253,243  
; FILING DATE: 04-OCT-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27,386  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (608) 283-2478  
; TELEFAX: (608) 251-5139  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2415 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Gossypium barbadense  
; STRAIN: Sea Island  
; IMMEDIATE SOURCE:  
; LIBRARY: EMBL-SI  
; CLONE: SIB12  
US-08-298-829-26

Query Match 9.2%; Score 193.8; DB 1; Length 2415;  
Best Local Similarity 99.0%; Pred. No. 4.3e-35;  
Matches 195; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agtgcgcgaacgaattaatgtagctcactcattagcaccacccaggctttacac 60  
Db 241 AGTGAGCGCAACGAATTAATGTAGCTAGCTCACTTAGCGACCCCGGCTTTACAC 300

QY 61 ttatgctccgcgctctatgtgtggaattgtgagcgatacaaatctcacacagga 120  
Db 301 TTTATGCTTCCGGCTCGTATGTTGTGGAATTTGAGCGGATAACAATTTACACAGGA 360  
QY 121 aacagctatgaccatgattacgcgaagctcgaaattaacccctcactaaagggaacaaaag 180  
Db 361 AACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCCTCACTAAAGGGAACAAAAG 420  
QY 181 ctggagctccacgcggt 197  
Db 421 CTGGAGCTCCACATGCT 437  
RESULT 15  
PCT-US96-06053-1/c  
; Sequence 1, Application PC/TUS9606053  
; GENERAL INFORMATION:  
; APPLICANT: Yale University  
; TITLE OF INVENTION: RECOMBINANT VESICULOVIRUSES AND THEIR  
; TITLE OF INVENTION: USES  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/06053  
; FILING DATE: 01-MAY-1996  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie  
; REGISTRATION NUMBER: 18,972  
; REFERENCE/DOCKET NUMBER: 6523-009-228  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14311 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 760..2028  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2092..2889  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2946..3635  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3774..5309  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 5429..11758  
PCT-US96-06053-1

Query Match 9.1%; Score 192; DB 6; Length 14311;  
Best Local Similarity 100.0%; Pred. No. 1.8e-34;

Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 agtgagcgcaacgcaattaatgtgagtttagctcactcattagcaccgccaggctttacac 60  
Db 12296 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACAC 12237  
Qy 61 ttatgtcttcggtcgtatgtgtgtggaattgtgagcgagataaacaatttcacacagga 120  
Db 12236 TTTATGCTTCGGCTCGTATGTTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 12177  
Qy 121 aacagctatgacctgattacgccaagctcgaaattaacccctcactaaagggaacaaaag 180  
Db 12176 AACAGCTATGACCATGATTAGCGCAAGCTCGAAATTAACCCCTCACTAAAGGGAACAAAAG 12117  
Qy 181 ctggagctccac 192  
Db 12116 CTGGAGCTCCAC 12105

Search completed: August 1, 2000, 18:20:49  
Job time: 74498 sec

2000-08-01 15:19:21

**THIS PAGE BLANK (USPTO)**







CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF101972; AAD04166.1; -  
DR PFAM; PF00201; UDPGT; 1.  
DR PROSITE; PS00375; UDPGT; 1.  
KW Transferase.  
SQ SEQUENCE 459 AA; 51411 MW; D7BBA527B61B6D08 CRC64;

Query Match 17.68; Score 435; DB 1; Length 459;  
Best Local Similarity 28.44; Pred. No. 1.7e-25;  
Matches 142; Conservative 81; Mismatches 191; Indels 86; Gaps 21;

Qy 4 PH-----VILTFPAQGHINPALQFAKNLVKMGIEVTFSTSIYAQSRMDEKSIILNAPKGL 58  
Db 9 PHEQVQVLLIPFPAQGHINPALQFAKNLVKMGIEVTFSTSIYAQSRMDEKSIILNAPKGL 58  
Qy 59 NFIPFS-----DGFDEGDSKDPVYMS-QLRKCGSETVKKIILTCSENGQPITCL 109  
Db 68 HFHAFQVPPFVSPPPEDDPFSLIPSEASHLR---EPVKGKLLQSLSSQAKRVVVI 123  
Qy 110 LYSIFLPAWAEVAREVHIPALLWSQATILDIYFNHGYEKAMANESND-----P 161  
Db 124 NDSLW-----ASVAQDA-----ANISNVENYTFHSP--SAFNTSGDFEEMGKPP 166  
Qy 162 NWSIQPLGLLETRDLPFLPYGAKGSLRVALLPPEKELIDTDAETPKILVNTFDEL 221  
Db 167 VGDHFHPEFPLEGCIARAF-----KG-FRTAQYERK-FNNGDIYNTSVIEGPPVEL 218  
Qy 222 EPEALNAIEGYKFGYIGLIPSAF-----LGGNDPLDASFGDLFQNSNDYMEWLNKPN 276  
Db 219 -LELFNG--GKKWALGPENFLAVEKSDSIGFRHPC-----MEWLDKQEP 260  
Qy 277 SSVVYISFGSLMNPISQMEISKGLIDIGRPLFWIKENEK-----GKEENKLL--GC 329  
Db 261 SSVYISFGTTLALDEQIQIATLEQSKQKFIWLRREADKGDIFAGSEAKRYELPKGF 320  
Qy 330 IELEKIGKIV-PWCSQLEVLKHPSLGCFVSHCGWNSALESACGVPVAPPTDQMTN 388  
Db 321 EERVEGHLVVRDAPQLEILSHSSTGGFMHCGWNSCLESITWGVPIATWPMISDQPRN 380  
Qy 389 AKQVEDVWKGVRVR--INEDGVVSESEIKRCIELVMDGGEGKEELRNKAKKRELAREA 446  
Db 381 AVLTEVLKGLVVKDWAQRNSLVASVYVNGVRRRLMETKE-GDEMQRVRLKNAIHS 439  
Qy 447 VKEGSSHKNLKAFIDVAK 466  
Db 440 MDEGGVSHMENGSTAHISK 459

RESULT 4  
ID UF05\_MANES STANDARD; PRT; 487 AA.  
AC Q40287;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 5 (EC 2.4.1.91) (UDP-GLUCOSE  
DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 5).  
GN GT5 OR UGT73A5.  
OS Manihot esculenta (Cassava) (Manioc).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Rosidae; eucosids I; Malpighiales;  
OC Euphorbiaceae; Manihot.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-COTYLEDON;

RX MEDLINE; 95201291.  
RT Hughes J., Hughes M.A.;  
RT "Multiple secondary plant product UDP-glucose glucosyltransferase  
RT genes expressed in cassava (Manihot esculenta Crantz) cotyledons";  
RL DNA Seq. 5:41-49(1994).  
CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS  
CC GLUCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN  
CC PIGMENTS (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL  
CC 3-O-D-GLUCOSIDE.  
CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE  
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.  
CC -1- TISSUE SPECIFICITY: FAINTLY EXPRESSED IN COTYLEDONS.  
CC -1- DEVELOPMENTAL STAGE: RARE TRANSCRIPTS EXPRESSED IN COTYLEDON AND  
CC ROOTS DURING THE DEVELOPMENT.  
CC -1- SIMILARITY: BELONGS TO THE UDP-GLUCOSYLTRANSFERASE FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X77462; CAA54612.1; -  
DR PFAM; PF00201; UDPGT; 1.  
DR PROSITE; PS00375; UDPGT; 1.  
KW Transferase; Glucosyltransferase; Multigene family.  
SQ SEQUENCE 487 AA; 54381 MW; 0F5CA2EBC897F124 CRC64;

Query Match 17.58; Score 432.5; DB 1; Length 487;  
Best Local Similarity 26.24; Pred. No. 2.8e-25;  
Matches 134; Conservative 96; Mismatches 193; Indels 89; Gaps 18;

Qy 3 QPHVLTTFPAQGHINPALQFAKNLVKMGIEVTFSTSIY---AOSRMDKSLTNA---PK 56  
Db 9 KPHVLTSSPGLHLIPVLELGRIVTL---CNFDVTFVWGSOTSAEPPQVLSAMTPK 65  
Qy 57 GLNFIPTSDGDEGFDHSDKP-----VFYMSQLKCGSETVKKIILTCSENGQPITC 108  
Db 66 LCEIQLP---PPNISCLIDPEATVCTRLFLVLMREIRAFRAAVAL-----KRP-AA 115  
Qy 109 LYSIFLPAWAEVAREVHIPALLWSQATILDIYFNHGYEKAMANESNDPNWSIQLP 168  
Db 116 IIVDLFGTSELEVAKELGIAYVYIASNAWFLALTIY-VPLDKVEGEFEVLQKEPKIP 174  
Qy 169 GLPLETRDLPFL-----PYGAKGSLRVALLPPEKELIDTDAETPKILVNTFDELEP 223  
Db 175 GCRPVREEVDPMLDRTNQOYSEYFRGLIEP-----TADGILMNTWEALEP 222  
Qy 224 BALNAIEGYKFG-----IGPLIPSAFLGGNDPLDASFGDLFQNSNDYMEWLNK 274  
Db 223 TTFGALRDVKELGRVAKVPVFPPIGFLRRQAGPCG-----SNCELLDMLDQO 268  
Qy 275 PNSSVYISFGSLMNPISQMEISKGLIDIGRPLFWIKENEK-----NEKGEE 322  
Db 269 PKESVYVYSGSGTSLSEQMLEAWGLERSQORFIWVVRQPTVKGTDAEFFTQDGGAD 328  
Qy 323 ENKLL--GCIELEKIGKIVP-WCSQLEVLKHPSLGCFVSHCGWNSALESACGVPVVA 379  
Db 329 MSGTFPEGLTRIQNVGLVFPQWSPQIHMSHPVSGVFLSHCGWNSVLESITAGVPIITAW 388  
Qy 380 PQWTDQMTNAKQVEDVWKGVRVR---INEDGVVSESEIKRCIELVMDGGEGKEELRNA 436  
Db 389 PIYAEQRNATLTTE--ELGVAVRPKNLPKAEVYKREIERMIRIM-VDEGSGIRRV 445  
Qy 437 KKKWELAREAYKEGSSHKNLKAFIDVAKGF 468  
Db 446 RELKDSGKALNEGSGSPNYMSALGNWEKSN 477



```
QY 120 EVAREVHIPALLWSOPATILD--IYFNHGYEKAMANESNDPNWSIQLPGL--PLLET 175
DB 7 DLADDFGIPSYIFFASGGFLGFMLYVOKIHDENFNFIKDKDSTELIVSLVNP-PPT 65
QY 176 RDLPSFLPYGAKGSLRVALPFPKELIDTDAETTPKILVNTFDFLEPEALNAIEGYKEY 235
DB 66 RILPSSILNKERFGQLLAIAKFRQ-----AKGIIVNTFLESRAIESFKVPPY 116
QY 236 GIGLPLPSAFLGGNDPLDASFGGDLFQNSN-DYMWLNSKNSVYIYFSGSLMNPISQ 294
DB 117 HVGPI-----LDVKSOG-----RNTHPFIMQMDQPGSVVFLCFGSMGFSBDQ 162
QY 285 MEEISKGLIDIGRPLWIKE-----NEKKEENKKL--GCIEELEKIGKIVPWC 344
DB 163 LKEIAYALENSGHRFLWSIRPPPPDKTASPTDYEDPRDVLPEGLFERTVAVGKVIHAP 222
QY 345 QLEVLKHPSLGCFVSHCGWNSALESACGVPVVAFPQMTDQMTNAKQVEDVWVKSQVRVI 404
DB 223 QVAVLAHPAIGGFVSHCGWNSVLESIFWFGVPIATWPMYAEQFNA--FEMVVELGLGVEI 280
QY 405 N-----EDG-VVESEIEKRCIELVMDGGEELKRNKAKKELAREAVKGGSSHKNLK 458
DB 281 DMGYRKESGIIVNSDKIERAKRLM-----ENDEKRRKKVEMREKSKMALIDGGSSFSILG 337
QY 459 AFIDDVAKG 467
DB 338 DFIKAMEG 346

RESULT 7
UFOL_MAIZE
ID UFOL_MAIZE STANDARD; PRT; 471 AA.
AC P16166.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE 3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MCC ALLELE).
GN BZ1 OR UG71A1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RA Furtak D., Schiefelbein J.W., Johnston F., Nelson O.E. Jr.;
RT "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL Plant Mol. Biol. 11:473-481(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88284304.
RA Ralston E.J., English J.J., Dooner H.K.;
RT "Sequence of three bronze alleles of maize and correlation with the
RT genetic fine structure.";
RL Genetics 119:185-197(1988).
CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X13500; CAA31855.1; "
```

```
DR EMBL; X07940; CAA30761.1; -.
DR PIR; S01052; S01052.
DR PIR; S08324; S08324.
DR MAIZEDB; 13885; -.
DR PFAM; PF0201; UDPGT; 2.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 471 AA; 48769 MW; 6234FD59219AF534 CRC64;

Query Match 16.8%; Score 417; DB 1; Length 471;
Best Local Similarity 26.7%; Pred. NO. 4e-24;
Matches 130; Conservative 72; Mismatches 234; Indels 50; Gaps 13;

QY 4 PHVILTTTPAQCCHINPALQFAKNL---VKMGIEVTFESTSIYAQSMDEKSIILNA---P 55
DB 12 PHVAVVAFPFSSHAVALLSIARALAAAAPSGATLUSFLSTASSLAQLRKASASAGHGUP 71
QY 56 KGLNFIPEFSDGDEGFDHSDKDPVFMYSQLRKCGSET--VKKII--LTCSENGQPTICLLY 111
DB 72 GNLRFVEVPDGAFAA--EETVPVPRQMQLFMEAEAGGVKAWLEAARAAAGGARVTCVVG 129
QY 112 SIFLPWAAEVAEREVHIPSAALLMSQPATILDIYFNFHGYEKAMANESND-----PNWSI 165
DB 130 DAFVWPAADAAASAGAPVWVWTAASCAL-----LAHIRTDLALREDVGDQAAANRVDGLLI 184
QY 166 QLPGLPLLETRDLPSEFLPYGAKGSLRVALPFPKELIDTDAETTPKILVNTFDELEPEA 225
DB 185 SHPGLASYRVRDLPGVV---SGDFNVYVNLVHRMGQCLPRSAVAALTFPGLDPPD 240
QY 226 LNAIEGYKFGYIGLIPSAFLTGGNDPLDASFGGDLFQNSNDYMWLNSKNSVYIYFSG 285
DB 241 VTAAELAEILPNCVPEGPHYLLLAEDDADTAAPAD----PHGCLAWLGRQPARGVAYVSG 296
QY 286 SLMNPSISQMEHISKGLIDIGRPLWVIKENE-----KGKEENKLGICIELEKIGKIV 340
DB 297 TVACPRPDELRELAAGLEDGSAFFLWSLREDSWPHLPPLPGFLDRAAGTGS-----GLVV 349
QY 341 PWCSQLVILKHPSLGCFVSHCGWNSALESACGVPVVAFPQMTDQMTNAKQVEDVWVKSQV 400
DB 350 PWAPQVAVLRHPSVGAFTVHAGWSVLEGLSSGVPACRPFFGQDRMNAARSVAHWVGFA 409
QY 401 RVRIINEDGVVESEIEKRCIELVMDGGEELKRNKAKKELAREAVKGGSSHKNLKAF 460
DB 410 AF-----EGAMTSAGVATAVEELL--RGEAGARMRAKELQALVAEAFGFGGECRKNFDRF 464
QY 461 IDDVAK 466
DB 465 VEIVCR 470

RESULT 8
UFOL_MAIZE
ID UFOL_MAIZE STANDARD; PRT; 471 AA.
AC P16167.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE 3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-W22 ALLELE).
GN BZ1 OR UG71A1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RA Furtak D., Schiefelbein J.W., Johnston F., Nelson O.E. Jr.;
RT "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL Plant Mol. Biol. 11:473-481(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88284304.
RA Ralston E.J., English J.J., Dooner H.K.;
```

\*Sequence of three bronze alleles of maize and correlation with the genetic fine structure.\*;  
 RL Genetics 119:185-197(1988).  
 CC -|- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS  
 CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN  
 CC PIGMENTS.  
 CC -|- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL  
 CC 3-O-D-GLUCOSIDE.  
 CC -|- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE  
 CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.  
 CC -|- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; X13502; CAA31857.1; -;  
 CC EMBL; X07937; CAA30760.1; -;  
 CC PIR; S01037; S01037.  
 CC PIR; S08326; S08326.  
 CC MAIZEDB; 13885; -;  
 CC PFAM; PF00201; UDPGT; 2.  
 CC PROSITE; PS00375; UDPGT; 1.  
 CC Transferase; Glycosyltransferase.  
 CC SW SEQUENCE 471 AA; 48673 MW; F417020B78366C01 CRC64;

Query Match	16.5%	Score 409;	DB 1;	Length 471;
Best Local Similarity	25.7%;	Pred. No. 1.6e-23;		
Matches	124;	Conservative 79;	Mismatches 236;	Indels 44; Gaps 13;
QY	4	PHVILTFPAQGHINPALQFAKNL----	VKMGIEVTFSTSIYAQSRMDEKSILNA----	P 55
DB				
DB	12	PHVAVAFPPSSHAUVLSTARALAAAAPSGATFLSFLSTASSLAQURKASSASAGHCLP	71	
QY	56	KGUNFIPFSDG---FDEGFDHSDKDPVFMYSQLRKCGETVKKIILTCSENGQPITCILLYS	112	
DB				
DB	72	GNLREVEVPDGAPEAEESVPPROMQLFM-EAAEAGGVKAWLEAARAAGGARVTCVVG	130	
QY	113	IFLPWAAEAVREWHIIPALLWSQPATLIDIIYFNPHGYEKAM-----ANESNDPNWSQLP	168	
DB				
DB	131	AFVWPADAASAAGAPVWPVTAAASCAL-LAHITRDALREDVGDOAANRVDEP--LISHP	187	
QY	169	GLPILLETDRLPSPLLPYGAGSLRVALPPPKELIDTLDAETTPKILVNTDELEPEALNA	228	
DB				
DB	188	GLASYRVDRIDPGVW-----SGDFNYVINLVLHRMGQCLPRSAAAVALNTFPGLDPPDVTA	243	
QY	229	IEGKVPYIGIPLPSAPFLGGNDPLDASFGGDLFQNSNDYMEWLNSKPNSSVVIYISFGSLM	288	
DB				
DB	244	ALAEILPNCVPFGPYHLLLAEDDADTAAPAD----	PHGCLAWLGRQARGVAVYSFGTVA 299	
QY	289	NPSISQMEETS KGLDIGRPFPLWIKENE-----KGKEENKKLGCIIELEKIGKIYPWC	343	
DB				
DB	300	CPRPDELRELAAGLEASGAPFLNSLRSDSWTLTLPFGFLDRAAGTGS-----GLVVYPWA	352	
QY	344	SOLEVLKHPSLGCFVSHCGHNSALESACGPVPVAPQWTDQMTNAKQVEDVWKSGYRVR	403	
DB				
DB	353	PQAVLURHPISGVAFVTHAGWASVLEGVSSGVPACRPFPGDQRNARSARVHMGFGAA	411	
QY	404	INEDGVYSESEIKRCITELVMDGKGEBELRKNKWKLEAREAVKEGSGSHKNLKAFID	463	
DB				
DB	412	---EGAMTSAGVAAVAEELL-REEEGARMARAKVLOALVAEAFGCGCEGRKNFDRVEI	467	
QY	464	VAK	466	
DB				
DB	468	VCR	470	

**RESULT 9**

```

UF0G_SOLME          STANDARD;          PRT;    433 AA.
AC   Q43641;
AC   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE).
DE   GT OR UGT76.
GN   OS
OS   Solanum melongena (Eggplant) (Aubergine).
OS   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
OC   Solanaceae; Solanum.
RN   [1]
RP   SEQUENCE FROM N.A.
RP   STRAIN=CV. SINGADOHARANASU; TISSUE=HYPOCOTYL;
RC   Toquri T.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC   -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC   GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC   PIGMENTS (BY SIMILARITY).
CC   -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC   3-O-D-GLUCOSIDE.
CC   -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC   ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
-----
CC   THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL; X77369; CAA54558.1; -.
DR   PFAM; PF00201; UDPGT; 1.
DR   PROSITE; PS00375; UDPGT; 1.
DR   Trnsterase; Glycosyltransferase.
SQ   SEQUENCE 433 AA; 48281 MW; 1EB76F9E5E63BBB CRC64;

```

```

Query Match      16.5%; Score 408.5; DB 1; Length 433;
Best Local Similarity 26.5%; Pred. No. 1.5e-23;
Matches 125; Conservative 84; Mismatches 184; Indels 79; Gaps 18;

QY 3 QPHVILTFPAQGHINPALQFAKNLVKMGIEVTEST--SIYAQSRMDEKSTLNAKPGLNPF 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 QLHIAFLAFPGTHATPLLTLVQ---KISFPLPSTIFSFTNTSSNSISFSPKVPNOENI 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 IPFS--DGFEQGFHSDKDPVYMSQLRKCQSETVKKIILTC-----SENGOPIT 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 KIYNVMDGVKRGND-----PPFGEALKLFIQSTLLTSKITEEAEEETGVKFS 109
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 108 CLLYSIFLPWA--AEVAREVHIPSALLWSOPATILDIYFNFHGYEKAMANESNDPNWSI 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110 CIFSDFAL-WCFVLKLPKKKMPAGVAYWTGSCSLAVHLYT---DLIRSNKET-----SL 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 166 QLPGL-PLLETRDLPSTLLPYGAGKSRVALPPEKELIDTDAETTPKILVNTDELEPE 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161 KIPGFSSTLSINDIPPEVTAEDLEGPMSSMLYNALNLRHADA-----VVLNSFQELD RD 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 225 AL-----NAIEGYPKYGTIGLPIPSAFLGNDPLDASFGGDLFQNSNDYMWELNSKPNS 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216 PLINKDLQKNL--QKVFENIGPLV---LOSSRKLDES-----GCIQWLMDKQEK 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 278 SVVYISFGSLMNPISOMEETSKGLIDIGRPFVLVIRENEKKEEENKKGCCIEELKIG 337
   ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 259 SVVYLSFGVTVTLPNPNEIGISAEALTKTKTFFIWSLRNN-----GVKKLPGLFRTKEFG 314
   ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 338 KIVPWCQLEKVLKHPISLGCYFSGHCWMSALESACGYVPVAVFPQWTDQMTNAKQVEDVWK 397
   ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 KIVSWAFPLETLAHKSVGVFTGHCWMSNILEGISFGVGPMTICRPFPGDOKLNSRVAVESWE 374
   ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

212 EL-LERPNG--GKEVWALGPFTPLAVEKKDSTGFSHPC-----MFWLDKQ 253

275 PNSSVYVIFSGSLMNPISIQMBEISIKSLIDIGRPFILWIKENEK-----GKEENKKL--327

254 EPSSVYVIFSGTTLALRDEQIQELATGLFQSKQKFIWLRLDADKGDIFDQSEAKRYELPE 313

328 GCTEELEKIGKIV-PWCSOLEVLKHPSLGCFVSHCGWNSALSACGVPVVAFFQWTDQM 386

314 GFEEVEGMLGVVRDAPQNEILSHSSTGTFMSHCGWNSCLESLSLGRGVPMATWAKHSDQP 373

387 TNAKQVEDYKSGVRVRINED--GVVESEEIKRCLTELVMDGKEEELRKNNAKKWKLAR 444

374 RNAVLTVDVLKGLIVKQWQKSLVSASVIEAVALRLMETRE-GDEIKRRAVLKLDKDIH 432

445 EAVKEGSGSHKLNKAFIDDDVAK 466

433 RSMDEGVSVMEMASFIHISR 454

RESULT 11

UF02\_MAIZE

ID UF02\_MAIZE STANDARD; PRT; 471 AA.

AC P16165;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID

DE 3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MC2 ALLELE).

GN B21 OR UGT71A1.

OS Zea mays (Maize).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.

[1]

RN SEQUENCE FROM N.A.

RA Furtak D., Schiefelbein J.W., Johnston F., Nelson O.E. Jr.,

RT "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";

RL Plant Mol. Biol. 11:473-481(1988).

CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS

CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN

CC PIGMENTS.

CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL

CC 3-O-D-GLUCOSIDE.

CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE

CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.

CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC

EMBL; X13501; CAA31856.1; -

DR PIR; S08325; S08325.

DR MAIZEDB; 13885; -

DR PFAM; PF00201; UDPGT; 2.

DR PROSITE; PS00375; UDPGT; 1.

KW Transferase; Glycosyltransferase.

SQ SEQUENCE 471 AA; 48621 MW; 81B897410A361299 CRC64;

Query Match 16.3%; Score 404; DB 1; Length 471;

Best Local Similarity 26.2%; Pred. No. 3.8e-23;

Matches 127; Conservative 77; Mismatches 234; Indels 46; Gaps 14;

4 PHYLITPFAQHINPALQFAKNL-----VKMGIEVTFSTSIYAQRMDKSLTNA-----p 55

12 PHAVAVAFPPSSHAALLSTARALAAAAPSGNTLSFLSTASSLAQLRKASSASAGHGLP 71

56 KGLNFIFFSDGDEGFDHSDKDPVFMYSQLRKCGSET--VKKII--LTCSENGQITCLLY 111

398 SGVVRINEDGVVESEEIKRCLTELVMDGKEEELRKNNAKKWKLAREAVKE 449

375 IGLQI---EGGIFTKGIISALDTFFN-BEKGKILRENVEGLKALEAVNQ 422

RESULT 10

ZOX\_PHAVU

ID ZOX\_PHAVU STANDARD; PRT; 454 AA.

AC P56725;

DT 15-FEB-2000 (Rel. 39, Created)

DT 15-FEB-2000 (Rel. 39, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)

DE ZEATIN O-XYLOSYLTRANSFERASE (EC 2.4.1.204) (ZEATIN O-BETA-D-

DE XYLOSYLTRANSFERASE).

GN ZOX1.

OS Phaseolus vulgaris (Kidney bean) (French bean).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; Rosidae; eurosids 1; Fabales; Fabaceae;

OC Papilionoideae; Phaseolus.

[1]

RN SEQUENCE FROM N.A.

RN STRAIN-CV, GREAT NORTHERN;

RX MEDLINE; 9292951.

RA Martin R.C., Mok M.C., Mok D.W.S.;

RT "A gene encoding the cytokinin enzyme zeatin O-xylosyltransferase of

RT Phaseolus vulgaris.";

RL Plant Physiol. 120:553-558(1999).

CC -1- FUNCTION: UTILIZES UDP-XYLOSE AS THE SUGAR DONOR AND CATALYZES THE

CC FORMATION OF O-XYLOSYLZEATIN FROM ZEATIN. DOES NOT ACT ON UDP-

CC GLUCOSE.

CC -1- CATALYTIC ACTIVITY: UDP-D-XYLOSE + ZEATIN = UDP + O-BETA-D-

CC XYLOSYLZEATIN.

CC -1- TISSUE SPECIFICITY: HIGH LEVEL IN YOUNG SEEDS, LESS IN OLDER SEEDS

CC AND VERY LOW IN ROOTS.

CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC

EMBL; AF116858; AAD51778.1; -

DR PROSITE; PS00375; UDPGT; 1.

SQ SEQUENCE 454 AA; 50972 MW; C57FCB6B34370A3E CRC64;

Query Match 16.3%; Score 404.5; DB 1; Length 454;

Best Local Similarity 26.7%; Pred. No. 3.3e-23;

Matches 134; Conservative 82; Mismatches 187; Indels 99; Gaps 19;

6 VILTTTFAQHINPALQFAKNLVMGIEVTFSTSIYAQRMDKSLNAPKGLNFIFFS- 64

11 VLLLPFPVQGLHLPFLQLSHLTAQAQVHYVGTV-THIRQAQLRYHYNATSIHFAPEV 69

65 -----DGFDEGFDHSDKDPVFMYS-QLRKCGSETVKKIIILTCSENGQITCLLYSIFLP 116

70 PPVVSPPNPDEDDFPSHLIPSFASAHLR-----EPVGKLLQLSLSSQAQRVVLINDSLMAS 125

117 WAAEAREVHIPSAALLWSQPATILDIYVFNHGYEK-----AMANESND-----PN 162

126 VAQDAA-----NFNVERVCFQVFSALNTAGDFWEQMGKPEL 162

163 WSIQLFGLPLLE---TRDLPSPLLPYTGAKGSLRVLPFPFKELIDTLDAETTPKILVTFD 219

163 ADFFHDPDIPSLQGCISAQFTDEL-----TAQNEFRK-FNNGDIYNTSRVIEGPV 211

220 ELPEALNATEGYKFGIGLIPSAF-----LGGNDPLDASPGDLPFONSNDYMEWLNK 274

212 EL-LERPNG--GKEVWALGPFTPLAVEKKDSTGFSHPC-----MENLDKQ 253

275 PNSSVYVIFSGSLMNPISQMBEISIKSLIDIGRPFILWIKENEK-----GKEENKKL--327

254 EPSSVYVIFSGTTLALRDEQIQELATGLFQSKQKFIWLRLDADKGDIFDQSEAKRYELPE 313

328 GCTEELEKIGKIV-PWCSOLEVLKHPSLGCFVSHCGWNSALSACGVPVVAFFQWTDQM 386

314 GFEEVEGMLGVVRDAPQNEILSHSSTGTFMSHCGWNSCLESLSLGRGVPMTATWAKHSDQP 373

387 TNAKQVEDYKSGVRVRINED--GVVESEEIKRCIELYMDGGEKEELRKNKAKKWKELAR 444

374 RNAVLTVDVLKGLVYKQKSLVSASVIEANVRRLMETRE-GDEIKRRAVLKDKIEH 432

445 EAVKEGSGSHKLNKAFIDDDVAK 466

433 RSMDEGVSVMEMASFIHISR 454

RESULT 11

UF02\_MAIZE STANDARD; PRT; 471 AA.

AC P16165;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID

DE 3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MC2 ALLELE).

GN B21 OR UGT71A1.

OS Zea mays (Maize).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.

[1]

RN SEQUENCE FROM N.A.

RA Furtak D., Schiefelbein J.W., Johnston F., Nelson O.E. Jr.,

RT "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays."

RL Plant Mol. Biol. 11:473-481(1988).

CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS

CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN

CC PIGMENTS.

CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL

CC 3-O-D-GLUCOSIDE.

CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE

CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.

CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; X13501; CAA31856.1; -

DR PIR; S08325; S08325.

DR MAIZEDB; 13885; -

DR PFAM; PF00201; UDPGT; 2.

DR PROSITE; PS00375; UDPGT; 1.

KW Transferase; Glycosyltransferase.

SQ SEQUENCE 471 AA; 48621 MW; 81B897410A361299 CRC64;

Query Match 16.3%; Score 404; DB 1; Length 471;

Best Local Similarity 26.2%; Pred. No. 3.8e-23;

Matches 127; Conservative 77; Mismatches 234; Indels 46; Gaps 14;

4 PHYLITPFAQHINPALQFAKNL-----VKMGIEVTFSTSIYAQRMDKSLTNA-----p 55

12 PHAVAVAFPPSSHAALLSTARALAAAAPSGNTLSFLSTASSLAQLRKASSASAGHGLP 71

56 KGLNFIFFSDGDFGDHSDKDPVFMYSQLRKCGSET--VKITII--LTCSENGQITCLLY 111

398 SGVVRINEDGVVESEIEIKRCIELYMDGGEKEELRKNKAKKWKELAREAVKE 449

375 IGLQI---EGGIFTKGIISALDTFFN-BEKGKILRENVEGLKALEAVNQ 422

RESULT 10

ZOX1\_PHAVU STANDARD; PRT; 454 AA.

AC P56725;

DT 15-FEB-2000 (Rel. 39, Created)

DT 15-FEB-2000 (Rel. 39, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)

DE ZEATIN O-XYLOSYLTRANSFERASE (EC 2.4.1.204) (ZEATIN O-BETA-D-

DE XYLOSYLTRANSFERASE).

GN ZOX1.

OS Phaseolus vulgaris (Kidney bean) (French bean).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; Rosidae; eurosids 1; Fabales; Fabaceae;

OC Papilionoideae; Phaseolus.

[1]

RN SEQUENCE FROM N.A.

RN STRAIN-CV, GREAT NORTHERN;

RX MEDLINE; 9292951.

RA Martin R.C., Mok M.C., Mok D.W.S.;

RT "A gene encoding the cytokinin enzyme zeatin O-xylosyltransferase of

RT Phaseolus vulgaris."

RL Plant Physiol. 120:553-558(1999).

CC -1- FUNCTION: UTILIZES UDP-XYLOSE AS THE SUGAR DONOR AND CATALYZES THE

CC FORMATION OF O-XYLOSYLEZATIN FROM ZEATIN. DOES NOT ACT ON UDP-

CC GLUCOSE.

CC -1- CATALYTIC ACTIVITY: UDP-D-XYLOSE + ZEATIN = UDP + O-BETA-D-

CC XYLOSYLEZATIN.

CC -1- TISSUE SPECIFICITY: HIGH LEVEL IN YOUNG SEEDS, LESS IN OLDER SEEDS

CC AND VERY LOW IN ROOTS.

CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; AF116858; AAD51778.1; -

DR PROSITE; PS00375; UDPGT; 1.

SQ SEQUENCE 454 AA; 50972 MW; C57FCB6B34370A3E CRC64;

Query Match 16.3%; Score 404.5; DB 1; Length 454;

Best Local Similarity 26.7%; Pred. No. 3.3e-23;

Matches 134; Conservative 82; Mismatches 187; Indels 99; Gaps 19;

6 VILTTTFAQHINPALQFAKNLVMGIEVTFSTSIYAQRMDKSLNAPKGLNFIFFS- 64

11 VLLLPFPVQGLHLPFLQLSHLTAQAQVHYVGTV-THIRQAQLRYHYNATSIHFAPEV 69

65 -----DGFDEGFDHSDKDPVFMYS-QLRKCGSEFVKKILTCSENGQITCLLYSIFLP 116

70 PPVVSPPNPDEDFPSHLIIPSFASAHLR-----EPVGKLLQLSLSSQAQRVVLINDSLMAS 125

117 WAAEAREVHIPALLWSQPATILDIYVFNHGYEK-----AMANESND-----PN 162

126 VAQDAA-----NFNVERVCFQVFSALNTAGDFWEQMGKPEL 162

163 WSIQLFGLPLLE---TRDLPSPLLPYGAKGSLRVLPFFKELIDTLDAETTPKILVTFD 219

163 ADFFHDPDIPSLQGCISAQFTDEL-----TAQNEFRK-FNNGDIYNTSRVIEGPV 211

220 ELPEALNATEGYKFGIGLIPSAF-----LGGNDPLDASPGDLFONSNDYMEWLNK 274

Db 72 GNLRFVEVPDGPAA--EETVPVPROMQLEMEAAEAGGVKAMLEAARAAGGARVTCVVG 129  
 QY 112 SIFLPAAEAVAREVHIPSALLKSPATILDIYFNHGYEKAM---ANESNDPNKSIQL 167  
 Db 130 DAFVWPAADAASAGAPVWVWTAASCAL-LAHIRTDLSREDVGQAANKRVDEP--LISH 186  
 QY 168 PGLPLETELDPSFLPYGAKGSLRVALPFFKELIDTLDAETTPKILVNTFDELEPEALN 227  
 Db 187 PGLASVYVRDLPDGVV---SGDFNVISLLVHRMGCLPRSAVAALNTFFGLDPDVT 242  
 QY 228 AIEGYFYIGIPLIPSAFAGLGGNDPLDASGGDLFQNSNDYMWLNKPNSSVYVIFSGSL 287  
 Db 243 AALAEILNCVPGPGPYHLLAEDADTAAPD---PHGCLAWLGRQARGVAVVSFGTV 298  
 QY 288 MNPISOMEIEISKGLIDICRPFLLWIKENE---KGKEENKKGICIELEKIGKIPW 342  
 Db 299 ACPRDELRELAAGLSAAPPFWLSUREDSWTLPPGFLDRAAGTS-----GLVVPW 351  
 QY 343 CSQLEKLPKSPFCVSHGWSNLSACGVPVPAFPQWTDQMTNAKQVEDVYKSGVRV 402  
 Db 352 APOVALRHPVSGAEVTHAGWASVLEGSSGVPMACRPFEGDQRMNARSVAHWGFGA 411  
 QY 403 RINEDGWSEEEKRCIELVMDGEGEELKNAKKWELAREAVKEGGSSHKNKAFID 462  
 Db 412 ----EGAMTSAGVAAVEELL-RGEEGAGMARAKELQALVAEAFGPGGECRKNDFRVE 466  
 QY 463 DVAK 466  
 Db 467 IVCR 470

## RESULT 12

UFO7\_MANES  
 ID UFO7\_MANES STANDARD; PRT; 394 AA.  
 AC Q40288;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 6 (EC 2.4.1.91) (UDP-GLUCOSE  
 FLAVONOID 3-O-GLUCOSYLTRANSFERASE 6) (FRAGMENT).  
 GN GT6 OR UGT73A6.  
 OS Manihot esculenta (Cassava) (Manioc).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;  
 OC Euphorbiaceae; Manihot.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COTYLEDON;  
 RX MEDLINE; 95201291.  
 RA Hughes J., Hughes M.A.;  
 RT "Multiple secondary plant product UDP-glucose glucosyltransferase  
 genes expressed in cassava (Manihot esculenta Crantz) cotyledons.";  
 RL DNA Seq. 5:41-49(1994).  
 CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS  
 GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN  
 PIGMENTS (BY SIMILARITY). MAY BE INVOLVED IN GLYCOSYLATION OF  
 CC INSTABLE CYANOHYDRINS TO PRODUCE STABLE CYANOGLUCOSIDES.  
 CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL  
 3-O-D-GLUCOSIDE.  
 CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE  
 ANTHOCYANIN BIOSYNTHETIC PATHWAY.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN COTYLEDONS AND LEAVES.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED PRIMARILY IN DEVELOPING COTYLEDONS  
 (FROM EMERGING GREEN COTYLEDONS TO APPROXIMATELY 10 DAYS OLD);  
 CC VERY LOW LEVELS IN HYPOCOTYLS AND NO MEASURABLE EXPRESSION IN  
 CC ROOTS.  
 CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X77463; CAH54613.1; -.  
 DR PFAM; PF00201; UDPGT; 1.  
 DR PROSITE; PS00375; UDPGT; 1.  
 KW Transferase; Glycosyltransferase; Multigene family.  
 FT NON\_TER 1  
 SQ SEQUENCE 394 AA; 44264 MW; FD51CD974E5ADCD7 CRC64;

Query Match 15.9%; Score 393.5; DB 1; Length 394;  
 Best Local Similarity 29.6%; Pred. No. 1.8e-22;  
 Matches 118; Conservative 76; Mismatches 155; Indels 49; Gaps 14;  
 QY 90 ETVKKIILTCSENGQITCLLYSIFLPWAAEVAAREVHIIPSAALLWSQPATILDIYF--NF 147  
 Db 18 EAVSKL---TARSDSSLAGFVLDMFCTSMIDVAKELGVPIYIFFTSGAAPLGFVQLI 74  
 QY 148 HGYEKAMANESNDPNKSIOLPGLPLETRDLPSEFLPYGAKGSLRVALPFFKELIDTLDA 207  
 Db 75 HDEQDADLTQFKRDSAELSVPSL---ANSLPARVLP--ASMLVKDRFYAFIRIRGL-- 126  
 QY 208 ETPPKILVNTFDELEPEALNAIEGYK-----FYGIGLIPSAFAGLGGNDPLDASFGDLFQ 262  
 Db 127 REAKGIWVNTFMELESALNSLKDQSKIPPIYVPGPILK---LSNOE-----NDVGP 176  
 QY 263 NSNDYMWLNKPNSSVYISFGSLMNPISQMEELISKGLIDIGRPFLLWIKENKGRKE 321  
 Db 177 EGSEIIEIWLDDQPSVFLFCFSGMGGMDDQAKELACALEQSRHFLMSLRPPPKGI 236  
 QY 322 E-----ENKK---LGCIELEKIGKIVPWCQSOLEVLPKHPSLGCFVSHGWSNLSAC 372  
 Db 237 ETSYDVENLOEILPVGFSERTAGKGVVWAPQVALEHPAIGGFVSHGWSNLSILESWF 296  
 QY 373 GVPVPAFPQWTDQMTNAKQVEDVYKSGVRVINEGDGWSE-----ETKRCIELVMDGG 426  
 Db 297 SVPIATWPLYAEQOFNAFTM--VTGLGLAVEIKMDYKKESEIILSADDIERGICKVM--- 351  
 QY 427 EKGEELKNAKKWELAREAVKEGGSSHKNKAFIDV 464  
 Db 352 EHHSEIRKRVKMSDKSRKALMDDESSFWLDRLIEDV 389

## RESULT 13

UFO7\_MANES  
 ID UFO7\_MANES STANDARD; PRT; 287 AA.  
 AC Q40289;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 7 (EC 2.4.1.91) (UDP-GLUCOSE  
 FLAVONOID 3-O-GLUCOSYLTRANSFERASE 7) (FRAGMENT).  
 GN GT7 OR UGT73A7.  
 OS Manihot esculenta (Cassava) (Manioc).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;  
 OC Euphorbiaceae; Manihot.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COTYLEDON;  
 RX MEDLINE; 95201291.  
 RA Hughes J., Hughes M.A.;  
 RT "Multiple secondary plant product UDP-glucose glucosyltransferase  
 genes expressed in cassava (Manihot esculenta Crantz) cotyledons.";  
 RL DNA Seq. 5:41-49(1994).  
 CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS  
 GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN  
 PIGMENTS (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL  
 3-O-D-GLUCOSIDE.  
 CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE  
 ANTHOCYANIN BIOSYNTHETIC PATHWAY.







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 14:38:44 ; Search time 3616.02 Seconds  
(without alignments)  
2567.101 Million cell updates/sec

Title: US-09-147-955-11  
Perfect score: 2105  
Sequence: 1 agtgagcgcaacgcaattaa.....acaaaaaaaaaaaaaaaaaaaaa 2105

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 5247842 seqs, 2204914090 residues 104955684  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1:	em_est1:
2:	em_est2:
3:	em_est3:
4:	em_est4:
5:	em_est5:
6:	em_est6:
7:	em_est7:
8:	em_est8:
9:	em_est9:
10:	em_est10:
11:	em_est11:
12:	em_est12:
13:	em_est13:
14:	em_est14:
15:	em_est15:
16:	em_est16:
17:	em_est17:
18:	em_est18:
19:	em_est19:
20:	gb_est1:
21:	gb_est2:
22:	gb_est3:
23:	gb_est4:
24:	gb_est5:
25:	gb_est6:
26:	gb_est7:
27:	gb_est8:
28:	gb_est9:
29:	gb_est10:
30:	gb_est11:
31:	gb_est12:
32:	gb_est13:
33:	gb_est14:
34:	gb_est15:
35:	gb_est16:
36:	gb_est17:
37:	gb_est18:
38:	gb_est19:
39:	gb_est20:
40:	gb_est21:
41:	gb_est22:
42:	gb_est23:
43:	gb_est24:
44:	gb_est25:
45:	gb_est26:
46:	gb_est27:
47:	gb_est28:
48:	gb_est29:
49:	gb_est30:
50:	gb_est31:
51:	gb_est32:
52:	em_est20:
53:	em_est21:
54:	em_est22:
55:	em_est23:
56:	em_est24:
57:	em_est25:
58:	em_est26:
59:	gb_est33:
60:	gb_est34:
61:	gb_est35:
62:	gb_est36:
63:	gb_est37:
64:	gb_est38:
65:	em_est27:
66:	em_est28:
67:	em_est29:
68:	em_est30:
69:	gb_est39:
70:	gb_est40:
71:	gb_est41:
72:	gb_est42:
73:	gb_est43:
74:	gb_est44:
75:	em_est31:
76:	em_est32:
77:	em_est33:
78:	em_est34:
79:	gb_est45:
80:	gb_est46:
81:	gb_est47:
82:	em_est35:
83:	em_est36:
84:	em_est37:
85:	gb_est48:
86:	gb_est49:
87:	gb_est50:
88:	gb_est51:
89:	gb_est52:
90:	gb_est53:
91:	gb_est54:
92:	gb_est55:
93:	gb_gss1:
94:	gb_gss2:
95:	gb_gss3:
96:	gb_gss4:
97:	em_gss1:
98:	em_gss2:
99:	em_gss3:
100:	em_gss4:
101:	gb_gss5:
102:	gb_gss6:
103:	gb_gss7:
104:	gb_gss8:
105:	gb_gss9:
106:	em_gss5:
107:	em_gss6:
108:	em_gss7:
109:	em_gss8:
110:	em_gss9:
111:	em_gss10:
112:	em_gss11:
113:	gb_gss10:
114:	gb_gss11:
115:	em_gss12:
116:	gb_gss12:





Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU  
5 prime sequence.

FEATURES

source Location/Qualifiers  
1. .649

/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEDN21G12"  
/clone\_lib="tomato fruit red ripe, TAMU"  
/tissue\_type="pericarp"  
/dev\_stage="red ripe (7-20 days post-breaker)"  
/note="Vector: pBlueScript SK(-); Site\_1: EcoR1; Site\_2: Xho1; supplier: Giovannoni; Fruit were tagged at the breaker stage (first sign of lycopene accumulation on the blossom end of the fruit) and harvested 7 days post-breaker (fully red-ripe), 10 days post breaker, and 20 days post-breaker (over-ripe). 20 day fruit which showed external or internal signs of pathogenesis were discarded. Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

BASE COUNT 217 a 118 c 124 g 190 t  
ORIGIN  
Query Match 17.0%; Score 356.8; DB 72; Length 649;  
Best Local Similarity 75.2%; Pred. No. 9.5e-74;  
Matches 487; Conservative 0; Mismatches 152; Indels 9; Gaps 3;

QY 657 taactgctacttactccattcttcttcttggcagcagaggttagcacgtgaagttc 716  
DB 2 TCACGTCCTCGTCTACACTCTCTCTACCTTGGCAGCTGAGGTAGCGGTGAAGTCC 61  
QY 717 acatccctctctctcttcttggagtcacacagcaacaatatggacatatattctca 776  
DB 62 ATATCCATCCCGCTTACTGATGATTCACACAGCAACTGTGCTAGACATATACACTATT 121  
QY 777 actttcagatgataaaagctat---ggctaataatccaatgacaaattgttcca 833  
DB 122 ACTTCAATGGCTATGAGGATGAATGAAGTGTAGTAGTCAATGAATGACAAATGGAGTA 181  
QY 834 ttcaacttcccggtctccactactggaactgagatctctcttcttcttactctctt 893  
DB 182 TCCAATTGCCAAGGCTTCCATTACTAATAAAGCCAAAGATCTTCCATCTTTTAGTTTCAT 241  
QY 894 atggtgcaaaa---ggagctcttcgagtgacacttccaccatccaagaattgatagaca 950  
DB 242 CTAGCTCAAAAGATGATAAGTATAGTTTGTCTCTACCAACATTCAAAGAGCAACTAGACA 301  
QY 951 cattagatgctgaacacactcctcaagattcttgtgaatacatttgatgaatgagcctg 1010  
DB 302 CATTAGATGCTGAAGAAATCCNAAGTACTTGTGAATACATTTGATGATGATGAGCTAG 361  
QY 1011 aggcactcaatgcaattgaggttataagttttatggaattggacogttgattccctctg 1070  
DB 362 AGCAGCTCAAGCTATTGAAAAATACAAATTAATGGAAATGGACCATGATTCCTTCAT 421  
QY 1071 cttctctgggtggaaatgaccccttttagatgcttcatttgggtgagatctttttcaaaatt 1130  
DB 422 CATTCCTGGGTGGAAGAAATTCATTGGAAATCTTCATTGGTGTGATCTTTTTCAAAAGT 481  
QY 1131 caa---atgactatagggaattgtaaacctcaaacgcaaatcatcattgatttatatat 1187  
DB 482 CAAATGATGACTACATGGGAATGTTTAAACACAAAGCCCTAAATCATCAATGTTTATATCT 541  
QY 1188 cttttgggtgctaatgacatccatttagccaatggaggagatcatcaaaaggggttga 1247  
DB 542 CATTTGGGAGTCTATTGAATTTATCAAGAAACCAAAAGGAGGAGATTCGAAGGGTTGA 601  
QY 1248 tagacatagggaagccgttttttatgggtgatataaaagaataatgaaaaag 1295  
DB 602 TAGAGATCCAAAGCCATTCTTATGGTGAATAAGAGATCAAGAGATCAAGGAAG 649

RESULT 4

AI771830  
LOCUS  
DEFINITION

AI771830 536 bp mRNA EST 29-JUN-1999  
EST252930 tomato ovary, TAMU Lycopersicon esculentum cDNA clone  
cLED38G17, mRNA sequence.

ACCESSION

AI771830

VERSION

AI771830.1

KEYWORDS

EST.

SOURCE

tomato.

ORGANISM

Lycopersicon esculentum

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;

Solanaceae; Solanum; Lycopersicon.

REFERENCE

AUTHORS

Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,  
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,  
Fuji,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,  
Martin,G.B., Tanksley,S.D. and Giovannoni,J.

TITLE

Generation of ESTs from tomato carpel tissue

JOURNAL

Unpublished (1999)

COMMENT

On Mar 8, 1999 this sequence version replaced gi:4388463.  
Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU

FEATURES

Location/Qualifiers

1. .536

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db\_xref="taxon:4081"

/clone="cLED38G17"

/clone\_lib="tomato ovary, TAMU"

/tissue\_type="carpel"

/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"

/note="Vector: pBlueScript SK(-); Site\_1: EcoR1; Site\_2: Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."

BASE COUNT 187 a 66 c 138 g 145 t

ORIGIN

Query Match 15.9%; Score 335.6; DB 44; Length 536;  
Best Local Similarity 79.1%; Pred. No. 8.8e-69;  
Matches 427; Conservative 0; Mismatches 104; Indels 9; Gaps 2;

QY 1083 gaatgaccttagatgcttcatttggtggtgattctttttcaaaattcaa---atgact 1139

DB 3 GAAAGATCATTTGGATCTTCATTTGGTGTGATCTTTTCAAAAGTCAATGATGACT 62

QY 1140 atatggaatggttaaaactcaaaagcgaatcattcagttgtttatatatcttttggagtc 1199

DB 63 ACATGAATGTTTAAACACAAAGCCCTAAATCATCAATTTGTTTATCTCATTTGGAGTC 122

QY 1200 taatgacatcatatagccaatgagagagatatacaaaaggttgatagacataggaa 1259

DB 123 TATTGAATTTATCAAGAAACCAAAAGGAGGAGATTCCAAAGGGTTGATAGATCCAA 182

QY 1260 ggcctgttttatgggtgataaaagaaaaatgaaagcgaagagagagaataaaaaagc 1319

DB 183 GGCCATCTTATGCGGTATTAAGAGATCAAGACAGACAAAAGAGAGAGAAATTA--- 239

QY 1320 ttggttattggaatggaaaaaataggaaaaaatagttccattggtgttcacacttg 1379

DB 240 ---GTTGCATGATGGAATTAGAGAAGCAAGGGAATAATAGTACCATGTTGTTTCACACTTG 296



/clone\_lib="tomato ovary, TAMU"  
/tissue\_type="carpel"  
/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"  
/lab\_host="Xl1-Blue MRP"  
/note="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and  
directionally cloned cDNA in vector Lambda ZAP II with 5',  
and 3' ends located at the EcoRI and XhoI sites,  
respectively."  
BASE COUNT 196 a 82 c 114 g 154 t  
ORIGIN

Query Match 15.4%; Score 325; DB 45; Length 546;  
Best Local Similarity 79.2%; Pred. No. 2.7e-66;  
Matches 414; Conservative 0; Mismatches 100; Indels 9; Gaps 2;  
QY 913 tcgagttgcactccaccattcaagaattgatagacacattagatgctgaaaccactcc 972  
Db 30 TAGTTTGTCTTACCAACATTAAGAGCAACTAGACACATTAGATGGTGAAGAAAATCC 89  
QY 973 taagattctgtgaatcatattgataattagagcctgagcactcaatgcaattgaagg 1032  
Db 90 AAAGGTACTTGTGAATACATTTGATGCATTAGAGCTAGAGCCCAATCAAGCTATTGAAA 149  
QY 1033 ttataagttttatggaattgacgcttgattcctctctctctctctctctctctctct 1092  
Db 150 ATACATTTAATTTGGAATTTGGACCATTTGATTCCTTCATCATCTTTGGTGGGAAGATTC 209  
QY 1093 tttagatgcttcatttggtggtgatctttttcacaattcaaa----atgactatatgaa 1149  
Db 210 ATTGGAATCTTCAATTTGGTGTGATCTTTTCAAAAGTCAAAATGATGACATGGAATG 269  
QY 1150 gtaactcaaaagccaaatcatcagttgtttatatatatcttttgggagtcataatgactc 1209  
Db 270 GTTAAACAAAGCCCTAAATCATCAATTTGCTTATATCTCATTTGGGAGTCTATTGAAAT 329  
QY 1210 atctattagcgaatgagagagatatcaaaagggttgatagacataggaagccgcttttt 1369  
Db 330 ATCAAGAACCAAGAGAGAGATTCGAAAAGGTTGTATAGATCCAAAGGCCATCTTT 389  
QY 1270 atgggtgataaaagaaatgaaagagcaaggaagagaagaataaaagccttggtgat 1329  
Db 390 ATGGGTAATAGAGATCACCAAGAAAGAAAGAAAGAG-----AAATTCAGTTGCAT 443  
QY 1330 tgaagattggaataaataaggaaaatagttccactgggttcacaacttgaagtctaaa 1389  
Db 444 GATGGAATTAGAGAAGCAAGGGAAAATAGTACCATGGTGTCACAACTTGAAGTCCTGAC 503  
QY 1390 acatcatctttaggagtttttctctctctctctctctctctctctctctctctctctct 1432  
Db 504 ACATCATCTTTAGGATGCTTGTCTCGCACTGTGGATGGAAT 546

RESULT 7  
AW651280 658 bp mRNA EST 04-APR-2000  
LOCUS EST329734 tomato germinating seedlings, TAMU Lycopersicon  
DEFINITION esculentum cDNA clone cLE116K15 5', mRNA sequence.  
ACCESSION AW651280  
VERSION AW651280.1 GI:7412518  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;  
Solanaceae; Solanum; Lycopersicon.  
REFERENCE 1 (bases 1 to 658)  
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,  
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,  
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and  
 Tanksley,S.D.  
TITLE Generation of ESTs from germinating tomato seed

JOURNAL Unpublished (2000)  
COMMENT On Jan 6, 2000 this sequence version replaced gi:6677182.  
Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU  
5 prime sequence.

FEATURES  
Location/Qualifiers  
source  
1..658  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLE116K15"  
/clone\_lib="tomato germinating seedlings, TAMU"  
/tissue\_type="whole seedlings"  
/dev\_stage="7 days post imbibition"  
/note="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; 7 days post imbibition on water-agar. Mixed stage  
whole germinating seedlings from seed coat emergence up  
to two centimeters in seeds not showing obvious signs of  
germination were discarded."  
BASE COUNT 208 a 126 c 124 g 200 t  
ORIGIN

Query Match 15.0%; Score 315.4; DB 79; Length 658;  
Best Local Similarity 71.7%; Pred. No. 5.2e-64;  
Matches 471; Conservative 0; Mismatches 176; Indels 10; Gaps 4;  
QY 566 gaccctgtatttcacatgtcacaaacttcgtaaatgtggaagtgaactgtcaaaaaata 625  
Db 2 GATCTAAACCTGATACATGTGAGAGATAAGAAAGTCGTGGGTCCCAACATTTGAGGGATGC 61  
QY 626 attctcacttgctgtaaaatggacagcctataactgctactcttactcattcttctt 685  
Db 62 ATTTTGAAGAGTTTCAGACGAGGACGCTCTGTCACTCCCTCGCTACACTCTCTCTGCTA 121  
QY 686 ccttgggcagcagaggttagcacgtgaagtccactcctctctctctctctctctctctct 745  
Db 122 CCTTGGCAGCTGAGGTAGCGCGTGAACCTCATATCCATCCCGCTTACTATGAATCAA 181  
QY 746 ccagcacaatatggacatatattacttcaactttctatgatatgaaagctat---g 802  
Db 182 CCAGCAACTGTGTAGACATATACTATTACTTCAATGGCTATGAGATGAATGAAG 241  
QY 803 gctaaatgaatccaatgatccaaattggtccattcaactcccggtctccactactgaa 862  
Db 242 TGTACTACCTCAATGATCCAAATTTGGAGTATCCAATTTGCCAAGGCTTCCATTACTAAA 301  
QY 863 actcgagatcttctctc-attttacttctctatggtgcaaaaggaggt---cttcaggt 918  
Db 302 AGCCAAGATCTTCATCTCTTTTGTAGTTTCATCTAGCTCAAAAGATGATGATAGTATGTT 361  
QY 919 tgcacttccaccattcaagaattgatagacacattagatgctgaaccactcctaagat 978  
Db 362 TGTCTACCAACATTAAGAGCAACTAGACACATTAGATGGTGGAGAAAATCCAAAGGT 421  
QY 979 tcttggaatacacattgatgaattagagcctgagcactcaatgcaattgaaggttataa 1038  
Db 422 ACTTGTGAATACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 481  
QY 1039 gttttatggaattggacggttgattcctctctctctctctctctctctctctctctctct 1098  
Db 482 TTTAATTTGGAATTTGGACCATTTGATTCCTTCATCTCTTGGTGGAAAAGATTCATTGGA 541  
QY 1099 tgcctcatttgggtgagatcttttcaaaattcaa---atgactatatggaatggtttaa 1155  
Db 542 ATCTCATTTTGGTGGTATCTTTTCAAAAGTCAAAATGATGATGATGATGATGATGATGAT 601  
QY 1156 ctcaagccaaattcactcagttgttttatatatcttttgggagtcataatgaatccatc 1212











/lab\_host="XLI-Blue MRF"  
/note="vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:  
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons  
of seedlings 7-10 days post-germination were excised, cut  
at both ends and placed on MS medium with no selection.  
Mixed callus was harvested at 25 and 40 days and included  
undifferentiated masses. Tomato Callus EST Library"  
BASE COUNT 177 a 142 c 137 g 194 t  
ORIGIN

Query Match 11.6%; Score 245.2; DB 63; Length 650;  
Best Local Similarity 65.3%; Pred. No. 1.7e-47;  
Matches 393; Conservative 0; Mismatches 203; Indels 6; Gaps 2;  
QY 310 aagtttttaattttcactaagaagaataatgtgagcagctcatgtcattcatttaacaac 369  
Db 24 AGGTTCTATTATTTAAGTTTGAGGTGAATAATGGTCAAGCCCATGTCTATTGGTGAC 83  
QY 370 atttcagcacaagggccatataatccagcagcttcaatttgccagaatctgttcaagat 429  
Db 84 ATTTTCTGCACAGGTCAATATCAATCCATCTGTTCAATTGGCCAGAGGCTAATCGAAAT 143  
QY 430 gggcatagaagtgacatttttacaagcatttatgcccagaagccggtatggatgaaatc 489  
Db 144 GGGCATTGAGGTGACATTACCACTAGTGTGTTTCCACCACCGTGTATGGCGAAATCGC 203  
QY 490 cattcttaagtacacaaaaggattgaatttcattccattttccgatggcttgatgaagg 549  
Db 204 GGCATTCCACGGCACCCCAAGGGCTTAAACTTGGCGGCATTTCTGATGGATTGATGAG 263  
QY 550 ttitga---tcattcaaaagccctgtattttacatgtcaccaacttcgttaaatgtgaag 606  
Db 264 TTTCAAGTCTAACGTGGATGATTTTAAACGTTACATGTACAGATACAGATGCTGGTGTC 323  
QY 607 tgaactgtcaaaaaataattctcacttgcctgtctgaaatggagagcctataaacttgct 666  
Db 324 CCAAACTATTGAGGTGTCATTTTGAAGAGTTCAGACGAGGAGCTTCTGTGCAGTCCCT 383  
QY 667 actttactccattttcttccttgggagcagaggtgagcagtgaaagtcaactcccttc 726  
Db 384 CGCTACACTCTTCTGTACCTTGGGAGCTGTATGACGCGTGAACCTCCATATCCCATC 443  
QY 727 tgcctctcttgagtgcaaccagcaacaatttgagacatatattacttcaacttctatgg 786  
Db 444 CGCGTTACTATGATTTCAACCGCACTGTGCTAGACATATACATCTATTTACTTCCATGG 503  
QY 787 atatgaaaaagctat---ggctaataatccaatgatcccaaatgggtccattcaacttcc 843  
Db 504 CTATGAGGATCAAAATGAAGTGTAGTCAATGATCAAAATGGAGTGTCCAAATGGCC 563  
QY 844 cgggcttcactactggaaactcagatcttccttcatttttacttccctatggtgaaaa 903  
Db 564 AAGGCTTCCATTACTATAAAACCAAGATCTTCCATCTTTTTTAGTTTATCTAGTCAAA 623  
QY 904 ag 905  
Db 624 AG 625

RESULT 15  
AW651250 642 bp mRNA EST 04-APR-2000  
LOCUS  
DEFINITION EST329704 tomato germinating seedlings, TAMU Lycopersicon  
esculentum cDNA clone cLEI116E21 5', mRNA sequence.  
ACCESSION  
VERSION AW651250  
KEYWORDS EST.  
SOURCE  
ORGANISM  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;  
Solanaceae; Solanum; Lycopersicon.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 642)  
Alicala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,  
Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Ronning, C.M.,  
Nierman, W., Fraser, C.M., Giovannoni, J.J., Martin, G.B. and  
Tankley, S.D.  
Generation of ESTs from germinating tomato seed  
Unpublished (2000)  
On Jan 6, 2000 this sequence version replaced gi:6677152.  
Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU  
5 prime sequence. Location/Qualifiers  
1. 642  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEI116E21"  
/clone\_lib="tomato germinating seedlings, TAMU"  
/tissue\_type="whole seedlings"  
/dev\_stage="7 days post imbibition"  
/note="vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:  
XhoI; 7 days post imbibition on water-agar. Mixed stage  
whole germinating seedlings from seed coat emergence up  
to two centimeters in seeds not showing obvious signs of  
germination were discarded."  
BASE COUNT 179 a 145 c 135 g 183 t  
ORIGIN

Query Match 11.3%; Score 238.4; DB 79; Length 642;  
Best Local Similarity 65.8%; Pred. No. 5.6e-46;  
Matches 379; Conservative 0; Mismatches 191; Indels 6; Gaps 2;  
QY 310 aagtttttaattttcactaagaagaataatgtgagcagctcatgtcattcatttaacaac 369  
Db 67 AGGTTCTATTATTTAAGTTTGTGAGTGTAAATAATGGTCAACCCCATGTCTATTGGTGAC 126  
QY 370 atttcagcacaagggccatataatccagcagcttcaatttgccagaatctgttcaagat 429  
Db 127 ATTTCTTCGCAAGAGTCAATATCAATCCATCTCTTCAATTGGCCAGAGGCTAATCGAAAT 186  
QY 430 gggcatagaagtgacatttttacaagcatttatgcccagaagccggtatggatgaaaaatc 489  
Db 187 GGGCATTGAGGTGACATTCACCACTAGTGTGTTTGGCCACCGTGTATGGCGAAATCGC 246  
QY 490 cattcttaagtacacaaaaggattgaatttcattccattttccagcttgcttgatgaagg 549  
Db 247 GGCATTCCACGGCACCCCAAGGGCTTAAACTTGGCGCATCTCTGATGGATTGATGATGG 306  
QY 550 ttitga---tcattcaaaagccctgtattttacatgtcaccaacttcgaaatgggaaag 606  
Db 307 TTTCAAGTCTAACGTGGATGATTTAAACGTTACATGTACAGATACAGATGCTGGTGTC 366  
QY 607 tgaactgtcaaaaaataatctcacttgcctgtgaaatggagcagcctataaacttgct 666  
Db 367 CCAAACTATTGAGGATGTCTATTTGAAGAGTTCAGACGAGGAGTCTCTGTGACGTCCCT 426  
QY 667 actttactccattttcttccttgggagcagaggtgagcagtgaaagtcaactcccttc 726  
Db 427 CGCTACACTCTTCTGTACCTTGGCAGCTGAGGTAGCGGTGAACCTCCATATATCCCATC 486  
QY 727 tgcctctcttgagtgcaaccagcaacaatttgagacatatattacttcaacttctatgg 786  
Db 487 CGCGTTACTATGATTTCAACCGCACTGTGCTAGACATATACATCTATTTACTTCAATGG 546  
QY 787 atatgaaaaagctat---ggctaataatccaatgatcccaaatgggtccattcaacttcc 843  
Db 547 CTATGAGGATCAAAATGAAGTGTAGTCAATGATCAATGATCAATTTGGAGTATCCCAATTGCC 606

Search completed: August 1, 2000, 14:38:55  
Job time: 62727 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:27:45 ; Search time 51.03 Seconds  
(without alignments)  
559.061 Million cell updates/sec

Title: US-09-147-955-6  
Perfect score: 2407  
Sequence: 1 MSRAHYLLATFPQAQHINPA.....EEDGSSVNNLKVFDEVGVI 461

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_64: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1054.5	43.8	456	2 C71402	probable glucosylt
2	1052	43.7	467	2 T02238	glucosyl transfera
3	935.5	38.9	458	2 C71420	hypothetical prote
4	751.5	31.2	471	2 A54739	indole-3-acetate b
5	740	30.7	460	2 T00639	hypothetical prote
6	711.5	29.6	479	2 E71419	probable indole-3-
7	697	29.0	456	2 T00506	indole-3-acetate b
8	682	28.3	438	2 T00507	indole-3-acetate b
9	682	28.3	484	2 D71419	probable indole-3-
10	656	27.3	453	2 T00511	indole-3-acetate b
11	631	26.2	475	2 T071419	probable indole-3-
12	504.5	21.0	455	2 T00584	indole-3-acetate b
13	483	20.1	440	2 T00583	indole-3-acetate b
14	482	20.0	453	2 T45603	glucosyltransferas
15	479	19.9	462	2 T01732	UTP-glucose glucos
16	457	19.0	476	2 T03745	glucosyltransferas
17	445	18.5	476	2 T03747	glucosyltransferas
18	443	18.4	451	2 T45604	glucosyltransferas
19	441	18.3	507	2 T46161	glucosyltransferas
20	431	17.9	347	2 T06371	probable UDP-glucu
21	429.5	17.8	433	2 S51767	glycosyl transfera
22	429	17.8	466	2 T07404	probable glucosylt
23	428	17.8	478	2 T05423	probable glucosylt
24	426.5	17.7	490	2 T46162	glucosyltransferas
25	424.5	17.6	449	2 T45605	glucosyltransferas
26	419.5	17.4	452	2 T12981	hypothetical prote
27	416.5	17.3	487	2 S41951	UTP-glucose glucos
28	412.5	17.1	438	2 T45602	glucosyltransferas
29	405.5	16.8	488	2 T07786	UDP-glucose glucos

30	399	16.6	447	2 T12978	hypothetical prote
31	393	16.3	455	1 XUBHEG	flavonol 3-O-gluc
32	389	16.2	452	2 T00981	flavonol 3-O-gluc
33	389	16.2	481	2 T01850	UTP-glucose glucos
34	386	16.0	478	2 A71417	hypothetical prote
35	383.5	15.9	478	2 T08395	UTP-glucose glucos
36	383	15.9	385	2 T12980	hypothetical prote
37	382	15.9	471	2 S01052	flavonol 3-O-gluc
38	379	15.7	420	2 T08005	flavonol 3-O-gluc
39	372.5	15.5	452	2 G71416	probable glucosylt
40	371.5	15.4	346	2 S41954	UTP-glucose glucos
41	366	15.2	471	2 S01037	flavonol 3-O-gluc
42	363.5	15.1	287	2 S41953	UTP-glucose glucos
43	362.5	15.1	472	2 S39507	glucuronosyl trans
44	361	15.0	449	2 S41950	UTP-glucose glucos
45	360	15.0	471	1 S08325	flavonol 3-O-gluc

ALIGNMENTS

RESULT 1

C71402

probable glucosyltransferase - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)

A:Variety: columbia

C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999

C:Accession: C71402

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D

P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; G

avanagh, T.; Hempel, S.; Kottler, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B

Nature 391, 485-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen

erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;

C.; Chalwatzis, N.

A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t

A:Reference number: A71400; MUID:98121113

A:Accession: C71402

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-456 <BEV>

A:Cross-references: GB:Z97335; NID:g2244747; PID:e327440; PID:g2244766

C:Genetics:

A:Map position: 4COP9-4G3845

C:Superfamily: flavonol 03-glucosyltransferase

Query Match 43.8%; Score 1054.5; DB 2; Length 456;  
Best Local Similarity 49.7%; Pred. No. 2.7e-73;  
Matches 232; Conservative 70; Mismatches 130; Indels 35; Gaps 14;

QY 3 RAHVLLATPPAQGHINPALQFAKRLANADIQVTFYTSVAMRRMSRTAGSGLINPVSF 62

Db 11 RPHYLLVTPPAQGHINPALQLANRLTHHGATVTYTAWSAHRMRGPPS--TKGL-SFAWF 68

QY 63 SGYDGGDLOPGDGKMYMSEKSRGKALKSDTLAAN-NVDKSSKITFVYVSHLFAWA 121

Db 69 TDGFDGDLKSFEDQKYMSELKRCGNSNALRDIKALDATTETEPITGVYVSLVPWVS 128

QY 122 VAREFHLRSGALLWIEPATVLDIDIFYFYFN-GYSDEIDAGSDAIHLPGGLPVLQORDPSFL 180

Db 129 VAREFHLPTLLWIEPATVLDIDIFYFYFN--EPIKLP-KLPITTTGDLPSFL 185

QY 181 LPSTHERFRSL-----MKEKLETEGEKPKVLVNSFDALPDALKADKYEMIAIGPL 234

Db 186 QP-----KALPSALVTLREHTEALETESNPKILVNTFSALEHDAITSVEKLMPIG 240

QY 235 IPFAFDLGDGDPDRSGGDLPEKGSNDCLWLSTNPSSVYVVSFGSVF-NITKSOME 293

Db 241 VSSS--EGKT-----DLFK--SSDEYTKWLSKLSRSVIYISLGHADDLPEKHME 288

QY 294 ETARGLLDCGRPFLVWVRVNEGEVLISCMEELKR---VGKIYVSWCSOLEVLTPSLGCF 350

Db 289 ALTHGVLTATNRFLWIVREKNPEKKNRFLLEIRGSDRGLVGVWCSQTAVALHCAVGC 348  
QY 351 VTHCGWNSTLESISFGVPMVAFPQWFOGTNAKLMEDEVWRTGVRVRANEGSVVDGDEIR 410  
Db 349 VTHCGWNSTLESISGVPMVAFPQWFOGTNAKLMEDEVWRTGVRVRANEGSVVDGDEIR 407  
QY 411 RCIEEVMGSGEAEEMRENAEKWKAMADAAEGGSPDLNLKGFVDE 454  
Db 408 RCLEVMGSGEAEEMRENAEKWKAMADAAEGGSPDLNLKGFVDE 454  
RESULT 2  
T02238  
glucosyl transferase, jasmonate-induced - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 24-Nov-1999  
C:Accession: T02238  
R:Kojima, H.; Hashizume, K.; Imanishi, S.; Nakamura, K.  
submitted to the EMBL Data Library, January 1997  
A:Description: Jasmonate-induced potential glucosyltransferase from tobacco suspension c  
A:Reference number: Z14633  
A:Accession: T02238  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-467 <KOJ>  
A:Cross-references: EMBL:AB000623; NID:d1094897; PID:d1019901  
A:Experimental source: strain BY-2  
C:Genetics:  
A:Gene: JIGT  
C:Superfamily: flavonol O3-glucosyltransferase  
Query Match 43.7%; Score 1052; DB 2; Length 467;  
Best Local Similarity 45.1%; Pred. No. 4.3e-73;  
Matches 211; Conservative 93; Mismatches 140; Indels 24; Gaps 11;  
QY 2 SRAHVLATPAQGHINPALQFAKRLANADIQVTFSTVYAWRMRSRTAAGSNGLINEFVS 61  
Db 6 NECHVLALPGGHHINPSLQFSKLLNLGVKVTLSLSLAFNRI-KNLPKIEGL-TFAP 63  
QY 62 FSDGYDDGLQPG-DDGKNYSEMKSRGKALSDTLAANNVDQKSK---TFVYVSHLFA 117  
Db 64 FSDGYDDGNFGSEDDYHLFNSAISKSHGSEFI-----ANLIKSAKNGYPTTRVYITLMD 118  
QY 118 WAAKVAAREHLRSALLWIEPATVLDIFVYFNGYSD---EIDAGSDAHLPGGLPVLQAR 174  
Db 119 WAGVAKLHIPSTLFIQIPATVFDIYYRTNFANFKYDSQDIIELP-GLPSLSS 177  
QY 175 DLPSFLLPSTHERFSL--MKEKLETEGEEKPKVLVNSFDALPDALKAIKDYEMIAIG 232  
Db 178 DFTSFVDDVKSNDWAVESIKRQIEILSENPRILVNTFDALNLRLVKNVTWVGIG 237  
QY 233 PLPSAFLDKSDPSRFGDGLPKGSDNDCLLEWLTNPRSSVYVYFSGFVNTTKSQM 292  
Db 238 PLPSSFLDEKDRDNFPAADMIB--SENINWELDARANKSVIYAFAGSYAIEISQWM 294  
QY 293 EETARGLLDGRPLLVVRV---NEGSEVLSCMEELKRVKIVSWCSOLEVTHPSLGC 349  
Db 295 EETSQGLLKGRPLVIRVETLNKEKEPEELTKCKDELEKIGRIVRWCSQMEVLKSHSGC 354  
QY 350 FVTHCGWNSTLESISFGVPMVAFPQWFOGTNAKLMEDEVWRTGVRVRANEGSVVDGDEI 409  
Db 355 FLTHCGWNSTLESISGVPVACPIWNDQICNAKLQIDVKNKIGVRVANKEG-IIRKDEF 413  
QY 410 RCIEEVMGSGEAEEMRENAEKWKAMADAAEGGSPDLNLKGFVDE 457  
Db 414 QRCIEVMGDAEEGELRKNAQKWKDLAKESTKNSNSSLNKLAYVNE 461  
RESULT 3  
C71420  
hypothetical protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)

A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999  
C:Accession: C71420  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; G  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;  
C.; Chalwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t  
A:Reference number: A71400; MUID:98121113  
A:Accession: C71420  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-458 <BEV>  
A:Cross-references: GB:Z97339; NID:g2244901; PID:e326935; PID:g2244911  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: flavonol O3-glucosyltransferase  
Query Match 38.9%; Score 935.5; DB 2; Length 458;  
Best Local Similarity 42.7%; Pred. No. 3.7e-64;  
Matches 206; Conservative 82; Mismatches 126; Indels 69; Gaps 16;  
QY 5 HVLATPAQGHINPALQFAKRLAN--ADIOVTFSTVYAW--RMSRTAAGSNGLINEFVS 61  
Db 13 HLEVTFPAQGHINPSLELAKRLAGTISGARVTFAAISAYNRRMFSTENVPETLI-FAT 71  
QY 62 FSDGYDDGLQPG-----DDGKNYSEMKSRGKALSDTLAANNVDQKSKTFVYVS 113  
Db 72 YSDGDDGFKSSAYSDKSRQDATCNFSEMRRRGKETLTIELIEDNR--KQNPFTCVVYT 129  
QY 114 HLFWAARVAREHLRSALLWIEPATVLDIFVYFNGYSDI-----DAGSDAHLPGGLP 169  
Db 130 ILLTWAEALA-----LFSIFVHFGYEDAISEMANTPSSSIKLP-SLP 172  
QY 170 VLAORDLPSTHERF--RSLMKEKLETEGEEKPKVLVNSFDALPDALKAI-DKYE 227  
Db 173 LLTVRDPISFVSNVYAFLLPAPREQIDSLKEINPKILINTFOELEPEAMSSVPDNFK 232  
QY 228 MIAIGTIPSAFLDGKPSDRSFGDGLPEKSDNDCLLEWLTNPRSSVYVYFSGFVNT 287  
Db 233 IVPVGPLL-----TLRTDPSRG-----EYEWLDTKADSSVLYSFGTLAVL 275  
QY 288 TKSQMEIARGLDCCGRPLVWV-----RVNEGEVLIISCEBELKRVKIVSWCS 337  
Db 276 SKQELVELCALIQSRRPFLWVITDKSYRNKDEQEKEEDCISSEFRELDEIGWVYSWCD 335  
QY 338 QLEVLTHPSLGCFTHCGWNSTLESISFGVPMVAFPQWFOGTNAKLMEDEVWRTGVRV-- 395  
Db 336 QFRVLNHRISGCFVTHCGWNSTLESISGVPMVAFPQWONDDMMNAKLLEDCWKTGVRVME 395  
QY 396 RANEEG-SVVDGIRRCIEEVMGSGEAEEMRENAEKWKAMADAAEGGSPDLNLKGFVDE 454  
Db 396 KKEEGVYVVDSEIRRCIEEVM--DKAEFRGNATRWKDLAAEAVREGSSFNHLKAF 453  
QY 455 LDE 457  
Db 454 VDE 456  
RESULT 4  
A54739  
Indole-3-acetate beta-glucosyltransferase (EC 2.4.1.121) - maize  
C:Species: Zea mays (maize)  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 29-Oct-1999  
C:Accession: A54739  
R:Szerszen, J.B.; Szczylowski, K.; Bandurski, R.S.  
Science 265, 1699-1701, 1994  
A:Title: Iaglu, a gene from Zea mays involved in conjugation of growth hormone indole  
A:Reference number: A54739; MUID:94367368



A:Accession: A54739  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-471 <SIZE>  
A:Cross-references: GB:I34847; NID:g548194; PIDN:AAA59054.1; PID:g548195  
C:Superfamily: flavonol O3-glucosyltransferase  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 31.2%; Score 751.5; DB 2; Length 471;  
Best Local Similarity 38.1%; Pred.No.5.le=50;  
Matches 182; Conservative 75; Mismatches 180; Indels 41; Gaps 13;

QY 5 HVLATPAQHINPQLAQFAKRLANADIQTFFTSVVAWRMSRTAAAGSGLINFVSFSD 64  
||| ||||| :||| :||| :||| :||| :||| :||| :|||  
DB 4 HLVPVPPGGHMPWQFAKLASKGVATTLVTT----RFIQTADVDAHPAMVEALSD 59  
:||| :||| :||| :||| :||| :||| :||| :|||  
QY 65 GYDDGLQGDDGKNYNSEMSESRGIKALSDTLAAANNVDOKSSK--ITFYVYSHLFWAAK 121  
:||| :||| :||| :||| :||| :||| :||| :|||  
DB 60 GHDEG---GFASAAGVAEYLEKQAASASIL-ASLVEARASSADAFCTCVYDSEDVWLPL 115  
:||| :||| :||| :||| :||| :||| :||| :|||  
QY 122 VAREFHLSALLWPETATVIDIFYFNG-----YSIEDAGSDAIHLPG---GLPV 170  
:||| :||| :||| :||| :||| :||| :||| :|||  
DB 116 VARRMGLPAPVFSTQCSCAVSAVVYHFSGRLAVPPGAADDGCGAGAALAASEAFLGLEPE 175  
:||| :||| :||| :||| :||| :||| :||| :|||  
QY 171 LAQRDLSPFLPSPTHERFRSLMKLELTLEECPK-VLVNSFDALPDALKADTKY-EM 228  
:||| :||| :||| :||| :||| :||| :||| :|||  
DB 176 MERSELFSFPV-DHGYPPTTMAQAIFAHAGKDDWLFNSFELETEVLAKITKYLUKA 233  
:||| :||| :||| :||| :||| :||| :||| :|||  
QY 229 IAGLPPIP-----SAFLDGKDPSDRSGGDLFEKGNDDCLEWMLTNPRSSVYYVSFG 282  
:||| :||| :||| :||| :||| :||| :||| :|||  
DB 234 RAIGPCVPPLTAGTAGANGR---ITYGANLVKP--EDACTKWLDTPKDRSVAYVSFG 286  
:||| :||| :||| :||| :||| :||| :||| :|||  
QY 283 SFVNNTKSOMEIEARGLLDCGPFLMWVRNNEGEVILISCMEELKRVG--KIWSVCSOLE 340  
:||| :||| :||| :||| :||| :||| :||| :|||  
DB 287 SLASLGNAQKEELARGLLIAAGKPFLLMVRASDEHQVPYLLAEATATGAAMVPMCPOLD 346  
:||| :||| :||| :||| :||| :||| :||| :|||  
QY 341 VLTHPSLCFCFTHCGWNSTLESISFGVMVAFPQWFQGTNAKLMEDYVRTGVRVRANE 400  
:||| :||| :||| :||| :||| :||| :||| :|||  
DB 347 VLAHPAVGCFTVTHCGWNSTLEALSFGVMVAMALTMDPTNARNVLAWGAVRRARDAG 406  
:||| :||| :||| :||| :||| :||| :||| :|||  
QY 401 GSVDYGDREIRCIEEVMDGGESRKLESAGKKDLARKAMEEDGSSVNNLKVFLDEV 458  
:||| :||| :||| :||| :||| :||| :||| :|||  
DB 407 AGVFLRGVERCKRVAMDGGEAASARKRAAGEWRDKRNNAVPGGSDRNLDEFVQV 464  
:||| :||| :||| :||| :||| :||| :||| :|||

RESULT 5  
T00639  
hypothetical protein F316.2 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999  
C:Accession: T00639  
R:Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo-João,  
Vysotskaya, V.S.; Yu, G.; Ecker, J.; Theodoridis, A.; Davis, R.W.  
submitted to the EMBL Data Library, February 1998  
A:Reference number: Z14197  
A:Accession: T00639  
A>Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-460 <FED>  
A:Cross-references: EMBL:AC002396; NID:g2749918; PID:g2829862; GSPDB:GN00059; ATSP:F316.2  
C:Genetics:  
A:Gene: ATSP:F316.2  
A:Map position: 1  
A:Introns: 219/3  
C:Superfamily: flavanol O3-glucosyltransferase

Query Match 30.7%; Score 740; DB 2; Length 460;  
Best Local Similarity 36.5%; Pred. No. 3.8e-49;  
Matches 173; Conservative 78; Mismatches 181; Indels 42; Gaps 13;  
Qy 3 RAHVLLATFPQAQHINPALQFAKFLANADIQVTFSTSYAWRRMSRTAAG-SNGLINFVS 61

[illegible]

Query Match	29.6%	Score 711.5;	DB 2;	Length 479;
Best Local Similarity	35.8%	Pred. No. 6.2e-47;		
Matches 172; Conservative	89;	Mismatches 174;	Indels 45;	Gaps 15;

Qy	5	HVLLATPPAQGHINPQLQFAKRLANADIQTFTTFSYAVRRMGRTRAGSN-----G	55
Db	8	HVMLVSPGGQGVNPLLRIKLITASKGLLYTFVTTEKPGWKQRANKIQDGVLPKPVGLG	67
Qy	56	LINFVSFSDGYDDGLQPDGDGNKYMSSEKSGIKALSDTLAANNVOOKSK-ITFYVYSH	114
Db	68	FIREFFSDFAD----DDEKRFDFAERPHLEAVCGKEIKNLVKRYNEPTVCLINNA	122
Qy	115	LFANAAKVAREFHURBSALLIEPATVIDIIFYFNQY-----SDEIDAGSDATHLPGGLP	169
Db	123	FVPWNCVDVAEELHIPSVAVLWSQCACILAYYYVHRHLVKEPTKETPEDI--SV EIP-CLP	178



C:Accession: D71419  
P.: Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dink  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans  
C.; Chalwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A:Reference number: A71400; MUID:98121113  
A:Accession: D71419  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-484 <BEV>  
A:Cross-references: GB:297339; NID:g2244901; PID:e327472; PID:g2244905  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 28.3%; Score 682; DB 2; Length 484;  
Best Local Similarity 35.5%; Pred. No. 1.2e-44;  
Matches 170; Conservative 92; Mismatches 173; Indels 44; Gaps 17;

QY 5 HVLATFPAQGHINPALQFAKRLANADIQVTFSTYAWRRMSRTA-----AGSN 54  
DB 13 HVLVYFQGHVNPVLLRLGLKLIASKLLVTVTT-ELWKKKRWQANKYDGLKPVGS- 70  
QY 55 GLNVFSDGVDGDDGKNTYMSKSGIKALSDTLAANNVDDKSKITPVYVSH 114  
DB 71 GSIRFEFFDEWAEDDRADPSLYIAHLESVIGREVSCLV--RRYEANEPVSCLINNP 128  
QY 115 LPAAKAVAREPHLSALLWIEPATVLDIFVYFNG--YSDEIDAGSDAHLPGGLPVL 171  
DB 129 FIPWVCHVAEENIFCAVLWVOSCAFSAYIHQDGSFPTPEPELD-VKPL-CVPLV 186  
QY 172 AQRDLPSFLPSTHERFRSLAKETLEGEKPK-VLVNSFDALPDALKAI-DKY-EMI 229  
DB 187 KNDIPLSFLRPSS--RFTGFRQAILQPKNLKSKFVLIDSDLEQEVYDYMSSICPVK 244  
QY 230 AIGPLIPSAFLDKGDPDRSGDGLFEKGSNDCCLEWLNTPRSVYVYVSGFVNTTK 289  
DB 245 TVGPLFKVARTVSDVS-----GDICK---STDCKLEWLDSPKSSVYVYSGTVAYLKQ 296  
QY 290 SOMEIARGLDGRFLWVR-----VNEGEVLISCHEELKRVGK--IVSWCSQLEVL 342  
DB 297 EQIEETAHGLVSLGSLFVLVIRPPHDLKVETHVLPQELKSSAGKGMIVDMCPQEQVL 356  
QY 343 THPSLCFVTHCGWNSTLESISFGVPMVAFQWFDQGTNAKLMEYVWRTGVRV-RANEEG 401  
DB 357 SHPSVACFVTHCGWNSTLESISFGVPMVAFQWFDQGTNAKLMEYVWRTGVRV-RANEEG 416  
QY 402 SVVDGDEIRRCIEVNDG--GBKSKRLRESAGKWKDLARKAMEEDGSSVNNLKVFLEV 458  
DB 417 RVVPREEV---AEKLEATVSGKAEELKRNALKWKAEEAAVAPGSSDKNFRFEVKL 472

RESULT 10  
T00511  
Indole-3-acetate beta-glucosyltransferase homolog T20D16.16 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999  
C:Accession: T00511  
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
submitted to the EMBL Data Library, November 1997  
A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.  
A:Reference number: 214159  
A:Accession: T00511  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-453 <ROU>  
A:Cross-references: EMBL:AC002391; NID:g2642427; PID:g2642442  
A:Experimental source: cultivar Columbia

C:Genetics:  
A:Map position: 2  
A:Introns: 266/2  
A:Note: T20D16.16  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 27.3%; Score 656; DB 2; Length 453;  
Best Local Similarity 34.9%; Pred. No. 1e-42;  
Matches 164; Conservative 88; Mismatches 164; Indels 54; Gaps 18;

QY 5 HVLATFPAQGHINPALQFAKRLANADIQVTFSTYAWRRMSRTAAGSNGLINVFSFD 64  
DB 10 HVLVYALPQGHLPMLKFAKHLARTNLHFTLAT-IESARDLLSSDDEPHSLVDLVFFSD 68  
QY 65 GYDDGLQGGDDGKNTYMSKSGIKALSDTL---AANNVDQ--KSKITFVYVSHLFAWA 119  
DB 69 GL-----PKDDPRDH-----EPLTESLRKVGANNFSKIIEGKRFDCCIISVPPTPW 114  
QY 120 AKVAREFHLSALLWIEPATVLDIFVYF---NGYSDEIDAGSDAHLPGGLPVLQORLD 176  
DB 115 PAVAAHNIPCAILMIEACAGFSVYRYRYTKNTNSFPD-LEDNPKVLEP-GLPFLVROL 172  
QY 177 PSFLPSTHERFRSLAKETLEGEKPKVLVNSFDALPDALKAI-DKYEMIAIGPLI 235  
DB 173 PTLMLPSHGAIFNTLMAEFVECL--KDKWVLANSFYELSVIESMFEDLKPIIPGLV 230  
QY 236 PSAFIDGKDPDRSGDGLFEKGSNDCCLEWLNTPRSVYVYVSGFVNTTKSOMEI 295  
DB 231 -SPFLIGAD-EDKILDKSLDMKADYCWELDKQV-----SILKSSNOVETI 278  
QY 296 ARGLLDCGRPFELWVVRNVEGE---VLISMEELKRVGKIVSWCSQLEVLTPSLGCFVT 352  
DB 279 ATALKNRGVPLFVLRPKKAENVDLEDMVEGQ--GVVIEGQEKILCHMAISCFVT 336  
QY 353 HCGWNSTLESISFGVPMVAFQWFDQGTNAKLMEYVWRTGVRV-RANEEGSVVDG---E 408  
DB 337 HCGWNSTLETVYVSGVPMVAFVTFDQPLDARLLVDVFGVGRMK-----NDVVDGELKVAE 392  
QY 409 IRRCTEYVMDGSEKSKRLRESAGKWKDLARKAMEEDGSSVNNLKVFLEV 458  
DB 393 VERCIDAVTKGTD-AADMRRAELKQATRSAMAPGGSILARNLDLFINDI 441

RESULT 11  
F71419  
probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 24-Nov-1999  
C:Accession: F71419  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; G  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;  
C.; Chalwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t  
A:Reference number: A71400; MUID:98121113  
A:Accession: F71419  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-475 <BEV>  
A:Cross-references: GB:297339; NID:g2244901; PID:e326932; PID:g2244907  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 26.2%; Score 631; DB 2; Length 475;  
Best Local Similarity 33.3%; Pred. No. 9.2e-41;  
Matches 165; Conservative 87; Mismatches 173; Indels 70; Gaps 19;

[illegible]

## 13

T00363  
indole-3-acetate beta-glucosyltransferase homolog T27E13.11 - Arabidopsis  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 11-Jan-2000  
C:Accession: T00583  
R:Roundley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syme  
submitted to the EMBL Data Library, May 1998  
A:Description: Arabidopsis thaliana chromosome II BAC T27E13 genomic sequen  
A:Reference number: 214177

## 12

A: Accession: T00383  
A: Status: translated from GE/EMBL/DDBJ  
A: Molecule type: DNA  
A: Residues: 1-440 <ROU>  
A: Cross-references: EMBL:AC004165; NID:g3150396; PID:g3150405  
A: Experimental source: cultivar Columbia  
C: Genetics:  
A: Map position: 2  
A: Introns: 149/1  
A: Note: T27E13.11  
C: Superfamily: flavonol O3-glucosyltransferase

-references: EMBL:AC004165: NID:q3150396: PID:q3150406

imental source: cultivar Columbia

ICS:

Position: 2

ns: 166/1

T27E13.12

Match	21 09.	Score	EM	E.	C.	reactions	155

Match

LOCAL SIMILARITY 30.0%, FREQ. NO. 4.3E-31;  
 Gaps 148: Conservative 83: Mismatches 178: Indels 75: Gaps 18:

5 HVLLATFPAOGHINPALOFAKRLAN--ADIOVTFTSVYAWRRMSRTAAGSNG-----I. 56

DATE	DESCRIPTION	AMOUNT	BALANCE
1/1/20	OPENING BALANCE		100.00
1/15/20	PAYROLL	50.00	50.00
1/20/20	RECEIVED	25.00	75.00
1/25/20	PAYROLL	50.00	25.00
1/30/20	RECEIVED	25.00	50.00
2/5/20	PAYROLL	50.00	0.00
2/10/20	RECEIVED	25.00	25.00
2/15/20	PAYROLL	50.00	0.00
2/20/20	RECEIVED	25.00	25.00
2/25/20	PAYROLL	50.00	0.00
2/28/20	RECEIVED	25.00	25.00
3/5/20	PAYROLL	50.00	0.00
3/10/20	RECEIVED	25.00	25.00
3/15/20	PAYROLL	50.00	0.00
3/20/20	RECEIVED	25.00	25.00
3/25/20	PAYROLL	50.00	0.00
3/30/20	RECEIVED	25.00	25.00
4/5/20	PAYROLL	50.00	0.00
4/10/20	RECEIVED	25.00	25.00
4/15/20	PAYROLL	50.00	0.00
4/20/20	RECEIVED	25.00	25.00
4/25/20	PAYROLL	50.00	0.00
4/30/20	RECEIVED	25.00	25.00
5/5/20	PAYROLL	50.00	0.00
5/10/20	RECEIVED	25.00	25.00
5/15/20	PAYROLL	50.00	0.00
5/20/20	RECEIVED	25.00	25.00
5/25/20	PAYROLL	50.00	0.00
5/30/20	RECEIVED	25.00	25.00
6/5/20	PAYROLL	50.00	0.00
6/10/20	RECEIVED	25.00	25.00
6/15/20	PAYROLL	50.00	0.00
6/20/20	RECEIVED	25.00	25.00
6/25/20	PAYROLL	50.00	0.00
6/30/20	RECEIVED	25.00	25.00
7/5/20	PAYROLL	50.00	0.00
7/10/20	RECEIVED	25.00	25.00
7/15/20	PAYROLL	50.00	0.00
7/20/20	RECEIVED	25.00	25.00
7/25/20	PAYROLL	50.00	0.00
7/30/20	RECEIVED	25.00	25.00
8/5/20	PAYROLL	50.00	0.00
8/10/20	RECEIVED	25.00	25.00
8/15/20	PAYROLL	50.00	0.00
8/20/20	RECEIVED	25.00	25.00
8/25/20	PAYROLL	50.00	0.00
8/30/20	RECEIVED	25.00	25.00
9/5/20	PAYROLL	50.00	0.00
9/10/20	RECEIVED	25.00	25.00
9/15/20	PAYROLL	50.00	0.00
9/20/20	RECEIVED	25.00	25.00
9/25/20	PAYROLL	50.00	0.00
9/30/20	RECEIVED	25.00	25.00
10/5/20	PAYROLL	50.00	0.00
10/10/20	RECEIVED	25.00	25.00
10/15/20	PAYROLL	50.00	0.00
10/20/20	RECEIVED	25.00	25.00
10/25/20	PAYROLL	50.00	0.00
10/30/20	RECEIVED	25.00	25.00
11/5/20	PAYROLL	50.00	0.00
11/10/20	RECEIVED	25.00	25.00
11/15/20	PAYROLL	50.00	0.00
11/20/20	RECEIVED	25.00	25.00
11/25/20	PAYROLL	50.00	0.00
11/30/20	RECEIVED	25.00	25.00
12/5/20	PAYROLL	50.00	0.00
12/10/20	RECEIVED	25.00	25.00
12/15/20	PAYROLL	50.00	0.00
12/20/20	RECEIVED	25.00	25.00
12/25/20	PAYROLL	50.00	0.00
12/30/20	RECEIVED	25.00	25.00
TOTAL			2500.00

13 HVVAMPYGRGHINPMMNLCKRLVRRYPNLHVTFFVT-EEW-----LGFIGDPKPKDR 64

57 INFVSFSDGYDDGLQPGDDGKNYMSEMKSRGIKALSDTLAANNVDQSSKITFVVYSHLF 110

[illegible]

63 IHFS'FLPNLIPSELVRAKDFIGFIDAVYTRL EEPFEKLL-----DSLNSppppSVIFADTY 11

```

Db      224  TGPLIPL-----EELSVGN-----NRELDYFKMLDEQPESVLYISGGFLSVSEA 270
      |||||
QY      291  QMEIEARGLLDCGRPFLVWVYVNEGEVLLISCEAEKLVKGIKVSQCLEVLTHPSLGCF 350
      |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      271  QMEETVGVVREANGVKFFWAR--GGELKLEALE--GSLGVVWVSWCDLRLVLCCHAAIGCF 326
      |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      351  VTHCGWNTLESISFGVPVMAFPQFMDQCTNAKLMEDEVRTGVRV-RANEEGSVVDGDEI 409
      |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      327  WTHCGYNTLEGICSGVPLLTFFVFDQFLNAKIVSEWRVGMGIERKKQMEILLIVSDEI 386
      |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      410  RRCLEEVMDG-GEKSKLRRESAGKKWDLARKAMEEDGSNNLKVFLDEVVGI 461
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      387  KELVKRFMDGSEGEKEMRRRTCDLSEICRGAVAKGSSDANIDAFIKDITKI 439
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT  14
T45603
glucosyltransferase-like protein - Arabidopsis thaliana
N;Alternate names: protein F12A12.180
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Feb-2000
C;Accession: T45603
R;Cholsne, N.; Robert, C.; Brottler, P.; Wincker, P.; Cattolico, L.; Artigou
submitted to the Protein Sequence Database, December 1999
A;Reference number: #23008
A;Accession: T45603
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-453 <CHO>
A;Cross-references: EMBL:AL139314
A;Experimental source: cultivar Columbia; BAC clone F12A12
C;Genetics:
A;Map position: 3
A;Introns: 161/1
A;Note: F12A12.180
C;Superfamily: flavonol O3-glucosyltransferase

```

```

Query Match      20.0%; Score 482; DB 2; Length 453;
Best Local Similarity 28.3%; Pred. No. 2.4e-29;
Matches 138; Conservative 85; Mismatches 185; Indels 80; Gaps 14;

QY      2  SRAHVLLATFPAGQHINPALQFAKRLANADIQVTFFTSYAWRRMSKRTAAGSNGLIINPVS 61
      :  :  :  ||||| :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      6  ARRSVVLVPFPAQGHISPMQLAKTLHLKGSFTVVQTKF-----NVFS 49

QY      62  FSDGVDGGLQ-----PGDGKN-----YSEMKSRGIKALSDTLAANNVQDKSKIT 108
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      50  PSDDFTHDFQVVTIPESLPESDFKNLGIQPLFKLNCKVSYFKDCL-GQLVLQGSNEIS 108

QY      109  FVWYSHLFAWAAKYAREHLRSALLWTEPATVLDIFYFYFNGYSDEIDA-----GSDAI 162
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      109  CVIYDEFWYFAAAAKCKLPNIIFTSTATACFASVFDKLVANNVQAPLKETGQQOE 168

QY      163  HLPGLGLVLAQROLPSLLPSTHERPESLMKEKLETEGEEKPKVLVNSPDALPEALKKA 222
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      169  LVPEYF-LAYKDF-----VSRFASLESIMEVYRTVDRKTASSVINTASCLESSLSF 223

QY      223  IDYEM-----IAIGPL-----IPSAFLDGKDPDSRFGGDLFEKGSNDDCDCLWLSTNPR 273
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      224  LQQQLQIPVYPIGLHNVASAPTSLE-----ENKSCIEWLNKQKV 265

QY      274  SSVVYVSFGSVNTTKSOMEETARGLLDGRPLWVVRVNE--GEVLISCEELKRV-- 329
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      266  NSVIYISMGSIALMEINMEIVASGLAASNOHFLWIRPGSIPGSEWIESMPEEFSKMWL 325

QY      330  --GKIYSCSQLEVLTHPSLGFVTHCGWNSTLESIFGVPVPAPFPQFQDQTNKLMED 387
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      326  DRGYIVKWAPOKEVLVSHPAQGVFWSHCGWNSTLESIGQVPMICRPFSCDQKVNARYLEC 385

QY      388  VVRTGVVRVANEEGSVVDGDETRRICEEVMDCGKSRKLRSAGKWAEDGGS 447
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

Search completed: August 1, 2000, 21:27:46

Job time: 11214 sec

